STATE NORMAL COLLEGE
YPSILANTI, MICH.

Vol. 33, No. 1 July, 1943

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

CONTENTS

Interrelations between Levels of Aspiration, Performance, and Estimates of Past Performance: J. A. BAYTON 1

Latency of Response in a Choice Discrimination: H. SCHLOSBERG AND R. L. SOLOMON 22

Remote Associations as a Function of the Length of Interval between Learning and Recall: J. T. WILSON 40

Generalization in the Initial Stages of Learning Nonsense Syllables: I. Integral Responses: B. R. PHILIP AND H. E. PEIXOTTO 50

Studies in Object-Preferences. I. The Effect of Temporal Proximity: F. W. IRWIN, F. M. ARMITT, AND C. W. SIMON 64

The Relationship between Attitudes toward Conclusions and Errors in Judging Logical Validity of Syllogisms: I. L. JANIS AND F. FRICK 73

The Retention of the Ability to Draw Lines of a Given Length Blindfold: E. O. BREGMAN, E. L. THORNDIKE, AND E. WOODYARD 78

PUBLISHED MONTHLY, TWO VOLUMES PER YEAR
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 665 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears monthly (since January, 1937), two volumes per year, each volume of six numbers containing about 520 pages.
Subscription: $14.00 ($7.00 per volume; Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 560 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, government, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

100
#6634

Journal of
Experimental Psychology

VOL. 33, NO. 1 JULY, 1943

INTERRELATIONS BETWEEN LEVELS OF ASPIRATION, PERFORMANCE, AND ESTIMATES OF PAST PERFORMANCE

BY JAMES A. BAYTON[1]

Virginia State College

INTRODUCTION

The level of aspiration experiment was conceived by Hoppe (17) as a technique for studying dynamic psychological factors which operate in the production of feelings of success and failure. Hoppe's experiments were characterized by a certain degree of informality, the conclusions being based upon the subjects' spontaneous remarks concerning their reactions to the various situations, the manner in which they went at the tasks, and their statements relative to success or failure. The rationale of his approach was that feelings of success or failure would be aroused by discrepancies between aspiration and subsequent performance. If the performance was equal to or exceeded the level of aspiration, success-feelings were presumed to be aroused, whereas if the performance fell below the aspiration, failure-feelings were presumed to result.

An important product of the research by Hoppe was a description of the needs which operate in the determination of a given level of aspiration. A level of aspiration was pictured as the resultant of three forces or needs. First was the need to keep the level of aspiration high. Later workers, notably Gould (11), have regarded this need as being an expression either of the individual's desire to do well or of his desire to appear to want to do well. Gould made this distinction on the basis of her discovery that some individuals give levels of aspiration which they do not think they will attain, but do so because they feel they are expected to appear ambitious. It is

[1] The author wishes to express his indebtedness to Dr. Malcolm G. Preston under whose immediate direction this experiment was done, and to acknowledge the benefit of discussions with Dr. Francis W. Irwin and the general encouragement of Dr. Samuel W. Fernberger.

obvious that this need, no matter what its origin, functions to elevate the level of aspiration. The second need comes from the desire to make the level of aspiration approximate the future performance as closely as possible. Frank (7) presents the view that this need operates to lower the level of aspiration. However, it is not difficult to conceive of two effects upon the level of aspiration of this need, one reflecting the subject's anticipation of improvement and the other his anxiety lest he fall victim to an accident and produce a poor performance. The direction of change in the level of aspiration which resulted from this particular need would depend, of course, upon the relative strengths of the two tensions involved. Finally, there is the need to avoid failure, which tends to exert a downward pull on the level of aspiration, since a greater possibility of failure exists when our aspirations are to a considerable degree in advance of our present level of performance.

The experimental work on the level of aspiration has been characterized for the most part by concern with the analysis of the sources, tension-systems, or needs from which the level of aspiration arises. Actually, however, the situation insofar as the individual is concerned does not terminate with the expression of his aspiration but continues into subsequent activities. If the needs described above have deeper significance than merely being determinants of aspiration, it would be expected that they would exert their influence upon the behavior *following* the statement of aspiration. These needs, as Hoppe (17) originally stated, are not satisfied merely by the statement of the aspiration, but only by what occurs after that. If the individual is to be certain of obtaining satisfaction and feelings of success, must he not *try* to make his performance equal or exceed his level of aspiration? Furthermore, is not the individual whose aspiration is more removed from his past performance level under greater tension to exert himself than is the individual whose aspiration is relatively close to his past performance level?

When we review the literature on the level of aspiration, we find that despite the fact that the term implies a subjectively determined goal toward which one is striving, there are few experiments which have tested the hypothesis that level of performance is in any way dependent upon the height of preceding aspiration. Filter (5) and Kneeland (19) obtained low correlations between estimates of future performance and performance. Yacorzynski (31) has reported that individuals exerting a minimum degree of effort had higher levels of aspiration than those exerting a maximum degree of effort. It should be noted that Yacorzynski did not concentrate upon performance values, being primarily interested in the quality of the behavior following the expression of the aspirations. He did refer to the per-

formance scores, reporting a tendency for them to be higher for the subjects putting forth the least effort, *i.e.*, those having the higher aspirations. In his experiment, however, this relationship must have been influenced by the fact that minimum degree of effort was indicated by the selection of less difficult methods of completing the tasks.

Notwithstanding this scarcity of experimental work on possible incentive functions of the level of aspiration, investigators have often suggested the possibility that such functions exist, some implying that the level of aspiration is an incentive and others questioning it. Hoppe in his use of the term level of aspiration meant that the subject undertook the task with 'certain demands' being made upon him. He then says that these 'expectations, goal-settings, or demands in connection with one's future performance' (**17**, p. 10) are expressed in the level of aspiration. Here it has been clearly implied that the level of aspiration is an expression of the manner in which one is approaching a task. Frank (**7**, p. 119) defined the level of aspiration as 'the level of future performance in a familiar task which an individual . . . *explicitly undertakes to reach.*' (Italics mine)

Gardner (**9**) has pointed to an important difference in the use of the term by Hoppe and Frank. Hoppe inferred the implicit level of aspiration from various statements of the subjects and observations of their behavior, whereas with Frank the level of aspiration was the explicit expression, or what the subjects *said* they were aspiring to attain. Subsequent research has followed Frank's procedure almost without exception.[2]

Gould (**12**) has referred to the level of aspiration technique (explicit) as a possible method for studying 'how much the future is weighted, how much of an upward pull it has . . .,' but concluded from interviews following her experiment (**11**) that we have yet to discover the nature of the relationship between implicit goal-strivings and the expressed level of aspiration. Rotter (**26**) and Gardner (**9**) have expressed this same doubt about the use of expressed level of

[2] Some investigators have questioned the key term used in the instructions given in those level of aspiration experiments which follow Frank's procedure. Irwin and Mintzer (**18**) have indicated the inconsistency which exists among previous workers, since some asked their subjects to state what they *expected* to make, one asked what his subjects *intended* to make, another asked "What *will* you make?", and still another used "What will you *try* to make?" The present experiment employs the most consistently used term—"expect." Irwin in personal communication with the writer has raised the question of whether instructions asking the subjects to state what they *expect* to make have the same meaning as those requiring them to state what they are *trying* to make, and claims that if the incentive value of the level of aspiration is being studied the latter instructions should be used. Since Irwin and Mintzer demonstrate a difference in aspiration statement when their instructions asked for *predictions* as against *hopes*, support is given Irwin's criticism. However, the present experiment is designed to test the incentive value of the level of aspiration as it is generally conceived, and it therefore uses the term most consistently found in the instructions of other investigators.

aspiration as a measure of the true amount of inner striving. Gardner asks:

> "Might not an individual in a task such as dart-throwing entertain at one and the same time a wild hope that he will make a perfect hit and a more prudent hope that he will at least hit the target, with perhaps an additional, self-conscious hope that he will not appear too awkward in the eyes of the experimenter? In other words, is there not considerable likelihood that an individual's aims on a given trial are manifold, fluctuant, ephemeral, and differing qualitatively as well as quantitatively, with those aims which involve a specific score often giving way to aims which cannot possibly be described in terms of score values?" (**9**, p. 65)

In summary it would seem that considerable doubt can be cast upon the claim that the level of aspiration statement is a measure of degree of self-motivation. One of the purposes of the present experiment is to test the hypothesis that performance is related to the relative height of the preceding explicit level of aspiration.

A feature of the experiments which pertains not only to method, but also to the analysis of the deeper structure of the personality forces underlying aspirations, has to do with the adequate control of the instructions so that the subjects can place consistent interpretations upon them, a problem which furnished the ground for the experiment by Irwin and Mintzer previously cited. Preston and Bayton (**24**) have defined three levels of aspiration. The Maximum level is that level which the subject feels represents his ultimate ability, the Actual level is the score the subject expects to make on the next trial, and the Least level is that score below which the subject is certain he will not fall. That these categories are of more than methodological significance is seen in the fact that they seem to serve at least two purposes for the individual. There is very substantial correlation between the Maximum and Actual levels, but the Least level has a low correlation with them (**25**). That the Least level is to a considerable extent independent of the Maximum and the Actual levels is also seen in the fact that a social variable tended to affect differentially the Least values as contrasted with those representing the Maximum and Actual levels (**24**). This differential character of the three levels has been further demonstrated by MacIntosh (**20**) under conditions which in all important particulars follow the procedure used by Preston and Bayton. Since it has been experimentally shown that these levels of aspiration have different meanings within personality, the important possibility exists that if there is dependence of performance upon aspiration statements it is a function of some particular level or set of levels.

Three considerations lead one to anticipate that any dependence which might exist would be a function of the Actual level of aspiration: (1) the Maximum level of aspiration is by definition removed from present performance in point of time; (2) the Least level is of *negative* valence as a score which the subject must not make; (3) the

Actual level is of immediate concern to the subject and reflects, perhaps, two tensions or needs—the need to do well and the need to avoid failure. Except for those few cases in which the Maximum and Actual levels coincide, it would not seem that the subjects should experience feelings of failure if the Maximum aspiration was not reached on the next trial. On the other hand, both the Actual and Least levels involve the possibilities of feelings of failure. But the Actual level is presumed to express also how well the individual feels he can do in the immediate future. Immediate satisfaction would seem to come not from avoiding the Least level but in attaining the Actual level of aspiration. The individual with the higher Actual level would seem to be under greater tension to make his performance match that aspiration since it is the one toward which attention is immediately directed.

We have been developing the theory that the needs which operate in the determination of the level of aspiration continue to exert their influence upon later performance. There is the further possibility that these needs are not dissipated upon completion of the performance *per se*, but extend into an evaluation of that performance. Before the subjects receive their scores from the experimenter there is evidence that they have tentative feelings of success or failure which are important determiners of the degree of satisfaction experienced. The present writer (**24**) has observed, in using fictitious scores, that some subjects claim that the score reported to them did not agree with their appreciation of the performance. They seemed to have made estimates of their performances; in fact, it is likely that these subjective estimates began to be formed during the performance. When the score was better than the subjective estimate surprise was expressed. However, most of these expressions came from those who felt their performances had been considerably better than the objective score reported to them.

If in the final analysis the needs are satisfied in terms of the appraisal of the performance, upon which of the preceding events, if any, is the estimate of past performance dependent? Its antecedents are the performance and the statement of the three levels of aspiration. In the present experiment the measure of performance ability used was scores based upon time, so that the estimate of past performance involved the ability to perceive the passage of time; and it has been shown that the reliability of such estimates is low (**29**). Kneeland (**19**) and Hilgard and Sait (**15**) have reported low correlations between estimates of past performance and the performances themselves.

If the theory presented above holds, *i.e.*, if the subject's definition of immediately doing well and avoiding failure is concentrated in the

Actual level, it would seem that the estimate of past performance would be dependent upon that particular level of aspiration. An individual may be expected to desire to obtain satisfaction out of his performance, if possible. Even when one feels that a given performance is poor, there still seems to exist a hope that the performance will turn out to be satisfactory. Therefore, the individual with the higher Actual level of aspiration would be expected to give the higher estimate of past performance, in view of this 'satisfaction-seeking' need which persists beyond the performance. It is the second purpose of this experiment to determine those factors in the level of aspiration situation upon which the estimate of past performance is dependent.

The factor of ego-involvement was recognized by Hoppe (**17**), and acknowledged by subsequent investigators, to play a critical role in the level of aspiration. The principle which has developed from this approach is that the tension represented in the needs is heightened in terms of the extent to which one's ego becomes involved in the task. Hoppe reported, for instance, that the level of aspiration had meaning only in those tasks which fitted the individual's level of accomplishment and which seemed to test his mettle, and which were not so easy as to seem silly or so difficult as to be impossible of solution. Kneeland (**19**) listed emotional attitude as a factor in the estimation of the future. Frank (**7**) found that an attitude of play affected the reliability of the statements of aspiration. Chapman and Volkmann (**3**), Gould and Lewis (**13**), Festinger (**4**), Hilgard, Sait, and Magaret (**16**), Preston and Bayton (**24**), and MacIntosh (**20**) have published results showing that the level of aspiration is a function of the individual's appreciation of his standing in terms of the performance and status of competing groups.

The evidence indicates that behavior in a level of aspiration situation is, in part, a function of the extent to which one feels his personality is involved in the performance of the task. It is possible that if there is dependence of performance upon aspiration it, too, is a function of ego-involvement. Perhaps levels of aspiration given in tasks taken lightly are not of incentive value, while an individual will be more concerned about the relation between his performance and his aspiration for that performance in a situation in which his ego-status is seriously involved. An individual taking a psychological examination which might mean the obtaining of a job must certainly approach that task more concerned about the relationship between his performance and his immediate task-goal than would a student taking the same examination as part of a laboratory exercise. In the same manner, two comparatively simple tasks can offer differential challenge to the individual's status. Of two tasks such as

arithmetic and cancellation one would expect that the average person would want to excel in the former, since the psychological importance of skill in cancellation is not immediately apparent but skill in mathematics is commonly accepted as indicating mental prowess. Since the strength of the tensions which exist in any level of aspiration experiment is a vital factor in the outcome, the present experiment includes within its design elements which will permit investigation of the influence of ego-involvement upon the extension of needs beyond the statement of the levels of aspiration.

Experimental Design

General Considerations.—The present experiment was designed to permit a study of the extent to which performance depended upon stated level of aspiration. Comparison of mean performance scores and mean aspirations (based on data from several trials) clearly would not answer this requirement, since such an analysis could not give us information on the cause-effect relationship. Hence, in using it we would be unable to discover whether the aspirations were determining the performances or the performances were determining the aspirations. To avoid this situation a trial-by-trial analysis was decided upon. By this method we could study each item (performance or estimate of performance) in terms of the events (chiefly aspirations) *immediately* preceding it.

The second requirement of the experimental design was the necessity of controlling variables, other than those relevant to the question under study, which might be influential in determining a difference under examination. The *S*s had to be matched in ability to perform the tasks before the introduction of the aspiration variables, since original differences between performers would tend to produce a spurious dependence of performance upon aspiration. In addition to the problem presented by the effect of true levels of performance upon the level of aspiration, there was the problem of the effect of subjective evaluation of that performance. Of two *S*s matched in ability to perform a task, one might *think* his last trial was superior and the other that his was inferior. This circumstance could influence the former to give higher levels of aspiration. In order to handle this problem the *S*s were matched on the basis of their estimates of past performance, wherever a study was made of the dependence of performance upon a level of aspiration. In addition, age and intelligence were inspected for possible influence upon our situation. Sex and race were controlled by selection of the *S*s. In order to exert some control on the region in which the various estimates would fluctuate, fictitious scores were presented to the *S*s in the early trials of the experiment.

Subjective involvement was determined by giving the *S*s brief experience with each of two tasks and asking them to designate the one in which a high degree of skill would be particularly satisfying. The task performed by the *S* in the experimental session had been previously determined by randomization, so that assignment to a task was not influenced by the choice made by the *S*. In this manner we had one group working on an ego-involved task and another on a non-ego-involved task.

Subjects.—All *S*s were Negro women attending Virginia State College. Each *S* was tested individually, the experimental session requiring approximately one hour. Three hundred *S*s were used, 150 for each task.

Tasks.—The two tasks were selected with the hope that subjective preference for a high degree of skill in one of them would be consistent. One of the tasks made use of arithmetic problems involving such items as the multiplication of four- and five-digit numbers, long division, and addition. The other task was a standard cancellation exercise.

Each *S* was given one trial on each task (Trials 1a and 1b). The *S*s were instructed to perform the t[illegible]s and at the end of these preliminary trials were asked, "Which one of these tasks would you prefer to excel in? In which one would the attainment of a high degree of skill be particularly satisfying?" The task indicated was then considered the ego-involved task. The tasks were alternated in presentation in order to prevent one from occupying a favored position. As we had hoped, all *S*s indicated that the arithmetic task was the one in which they would prefer to excel.

Actual Performance Level.—The first part of the experimental session was designed to ascertain the *S*'s level of performance. Following the preliminary trials, seven trials were given. For these trials, and all others throughout the experiment, the time in seconds taken to complete the task was recorded without the *S*s being aware of the true values. The *S*s were told that accuracy was essential and that this would be checked later.

Estimates of Past Performance.—Fictitious scores were given to each *S* after each of the above trials with the exception of the last one (after Trials 2 to 7, but not after Trial 8). After Trial 8 the *S*s, instead of being given a score, were asked to estimate what that score had been. In order to assist the *S*s in evaluating their performance (and later their levels of aspiration) the fictitious scores were presented before Trial 2 as falling in a scale ranging from 0 to 300. The experimenter said, "I am going to tell you your score after you do each trial. I will record the time it takes you but I will have to change it from seconds into scores for statistical reasons. If you don't finish the trial your score will be zero, 150 is the average obtained in other experiments, and a score of 300 is impossible, like running the 100-yard dash in five seconds." At the completion of Trial 8 the experimenter said, "This time you tell me what you think you made." Estimates of past performance were obtained after Trials 8, 9, and 10. No further performance information was given by the experimenter during these trials. Table 1 gives the prearranged scores used. They were the same for both tasks.

TABLE 1

PREARRANGED SCORES

Trial	Scores
2	146
3	148
4	147
5	149
6	149
7	151
8	—

Levels of Aspiration.—Immediately after the estimate of past performance for Trial 8 was obtained, the *S*s were instructed to give the Maximum, Actual and Least levels of aspiration for Trial 9. The experimenter was careful that no mention of anything pertaining to aspiration had entered the situation until this point. The instructions for the various levels were as follows: "Now I want you to tell me what you are going to do this next time. However, you will have to give me three different statements. First I want you to tell me the best score you think you could ever make in this task. Then I want you to tell me that score below which you are certain you will not fall if you should have a bad trial. Finally, tell me what you feel you will do this next time, exactly what you expect to make."

Critical Performances.—The above procedure (giving estimates of the previous performance, giving levels of aspiration, and performing the task) was given for Trials 9, 10, and 11. These trials, then, became the critical performances in the experiment.

Influence of Age and Intelligence.—It was necessary to inspect our data in order to determine whether age and intelligence would have to be taken into consideration in our various matching procedures. Table 2 gives the correlations between age and performance, age and levels of aspiration, intelligence and performance, and intelligence and levels of aspiration, for both tasks. These correlations were based upon the data of the 150 *S*s in each task. The measure of performance used was the sum of the time in seconds for Trials 6, 7, and 8, these trials being selected because they represented the *S*s' ability just prior to the introduction of the experimental conditions. The *r*'s between age and performance in cancellation and arithmetic were .22 ± .05 and .02 ± .05, respectively. In the case of intelligence, the measure being the IQ's made on entrance as Freshmen, the *r*'s were − .05 ± .05 and .13 ± .05, respectively. The low *r* between IQ and speed in cancellation is in keeping with those reported by others (10). The low *r* for IQ and arithmetic might seem surprising; it is accounted for by the fact that the criterion in the task was time and not accuracy. The problems were simple and, although the necessity for accuracy was stressed in the instructions, only the time in seconds was recorded. The aspiration data used in the correlations were those given for Trial 9. The table reveals that age and intelligence were not related to either the Maximum, Actual, or Least level of aspiration in either task.

TABLE 2

CORRELATIONS BETWEEN AGE AND PERFORMANCE, AGE AND LEVELS OF ASPIRATION, INTELLIGENCE AND PERFORMANCE, AND INTELLIGENCE AND LEVELS OF ASPIRATION

Variable	Age		IQ	
	Cancellation	Arithmetic	Cancellation	Arithmetic
Σ 6, 7, 8	.22±.05	.02±.05	−.05±.05	.13±.05
Maximum	.02±.05	.16±.05	.05±.05	.04±.05
Least	−.15±.05	.05±.05	.07±.05	−.01±.05
Actual	.00±.05	.09±.05	.05±.05	.01±.05

Because of these low *r*'s it was decided that age and intelligence could be eliminated from consideration in the subsequent analyses.

RESULTS

Dependence of Performance Upon the Levels of Aspiration.—In the analysis of the nature of the dependence of performance upon the levels of aspiration, the *S*s within a given task were matched for previous performance ability and subjective estimate of the performance just preceding the levels of aspiration. Table 3 gives the data showing that the pairs of *S*s were adequately matched. In the analysis of the influence of the Maximum level of aspiration in the cancellation task there were 44 pairs of *S*s simultaneously matched for previous performance (the sum of the three preceding performances) and estimate of past performance as these values were yielded just prior to Trial 9. The P-value of the mean of the difference between these pairs was between .36 and .35 in the case of the sum of the performances on Trials 6, 7, and 8. The P-value of the difference in variability between the two groups was .46. This indicated that any difference in performance ability which existed prior to the stating of the Maximum aspirations for Trial 9 was due to chance. The P-value was between .18 and .17 for the mean difference in estimate of the performance and the P-value of the difference in variability was between .24 and .23 at Trial 8, showing that there was also no reliable difference between the matched *S*s in their estimate of past performance prior to the statement of the Maximum aspirations for Trial 9. In cancellation there were 49 pairs of *S*s matched for the two variables in the analysis of the influence of difference in Maximum aspiration for Trial 10 upon performance at Trial 10. The P-value of the mean difference in performance at Trials 7, 8, and 9 was between .35 and .33, while the P-value of mean difference in estimate of what performance at Trial 9 had been was .44. The P-values of the differences in variability for the corresponding measures were between .20 and .21 and .48. In summary, in the analysis

TABLE 3

SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS MATCHED FOR PAST PERFORMANCE AND ESTIMATED PAST PERFORMANCE

Trial	Cancellation					Arithmetic				
	Number of Matched Pairs	P Mn. Diff. in Past Performance	P Diff. in σ's Past Performance	P Mn. Diff. in Est. Past Performance	P Diff. in σ's Est. Past Performance	Number of Matched Pairs	P Mn. Diff. in Past Performance	P Diff. in σ's Past Performance	P Mn. Diff. in Est. Past Performance	P Diff. in σ's Est. Past Performance
					Maximum					
9	44	.36>P>.35	.46	.18>P>.17	.24>P>.23	46	.30>P>.28	.38	.33	.40>P>.38
10	49	.35>P>.33	.21>P>.20	.44	.48	47	.48>P>.46	.50	.46>P>.44	.33>P>.31
11	45	.14>P>.13	.20	.11>P>.10	.21>P>.20	46	.40>P>.38	.36>P>.35	.50	.50
					Actual					
9	49	.33>P>.31	.46>P>.42	.33>P>.31	.31	51	.15>P>.14	.21>P>.20	.33>P>.31	.35
10	55	.42>P>.40	.27>P>.26	.31>P>.29	.40>P>.38	55	.38>P>.36	.40	.29	.40>P>.38
11	52	.10>P>.09	.29	.46>P>.42	.46	52	.44	.40>P>.38	.48>P>.46	.46>P>.42
					Least					
9	40	.24>P>.23	.31>P>.29	.13	.36>P>.35	43	.11>P>.10	.23>P>.21	.48>P>.46	.33>P>.31
10	47	.46>P>.44	.50>P>.48	.23>P>.21	.40	52	.44>P>.42	.35>P>.33	.40>P>.38	.46>P>.44
11	52	.14>P>.13	.30	.29>P>.27	.42	47	.46>P>.44	.50	.27>P>.26	.44

of the dependence of performance on Trial 10 upon the Maximum aspirations for that trial, the *Ss* were adequately matched for previous performance and estimate of previous performance. Inspection of Table 3 shows that on no trial in either task was there any reliable difference between the pairs of *Ss*, either in terms of central tendency or variability in previous performance and estimate of previous performance.

The dependence of performance upon the levels of aspiration is given in Table 4. A significant difference in the levels of aspiration was expected, according to the hypothesis of the experiment. In handling the performance scores the following procedure was used. Whenever the *S* in the High aspiration group gave a faster performance than her match in the Low aspiration group, that difference was

TABLE 4

DEPENDENCE OF PERFORMANCE UPON LEVELS OF ASPIRATION *

Trial	Cancellation			Arithmetic		
	P Mn. Diff. in Level of Aspiration	Mn. Diff. in Performance and P	% >0 and P	P Mn. Diff. in Level of Aspiration	Mn. Diff. in Performance and P	% >0 and P
			Maximum			
9	>.01	.00 (.50)	54 (.70>P>.50)	>.01	−3.92 (.16>P>.15)	48 (.80>P>.70)
10	>.01	−.26 (.40>P>.38)	53 (.70>P>.50)	>.01	−3.51 (.23>P>.21)	45 (.50>P>.30)
11	>.01	.11 (.46>P>.44)	47 (.70>P>.50)	>.01	−1.50 (.35>P>.33)	39 (.20>P>.10)
			Actual			
9	>.01	1.53 (.17>P>.16)	59 (.20>P>.10)	>.01	−3.82 (.14)	43 (.50>P>.30)
10	>.01	.35 (.35>P>.33)	54 (.70>P>.50)	>.01	−5.54 (.07>P>.06)	44 (.50>P>.30)
11	>.01	−.35 (.35>P>.33)	48 (.95>P>.90)	>.01	−4.42 (.09>P>.08)	50 (.99>P>.98)
			Least			
9	>.01	.75 (.35>P>.33)	60 (.30>P>.20)	>.01	−1.28 (.38>P>.36)	49 (.90>P>.80)
10	>.01	.88 (.09)	57 (.50>P>.30)	>.01	−.10 (.42>P>.40)	54 (.70>P>.50)
11	>.01	.39 (.31>P>.29)	56 (.50>P>.30)	>.01	−.53 (.44)	53 (.50>P>.30)

* *Negative* mean differences in performance indicate that the subject with the higher aspiration produced a faster performance than her match.

given a minus sign. Therefore, negative mean differences in performance which appear in the table show that the High aspiration group tended to give a faster performance than the Low aspiration group. Such negative differences would support the major hypothesis.

When the 44 pairs of *S*s matched for previous performance and estimate of past performance in cancellation (Table 3) were divided into groups having high and low Maximum levels of aspiration for Trial 9, the P-value of the mean of the difference in Maximum aspiration was less than .01, indicating that they had been reliably divided in Maximum aspiration. In cancellation the mean difference in performance on Trial 9 was .00 seconds with a P-value of .50, forcing the conclusion that the groups did not differ significantly in performance on Trial 9. An additional check on this performance was obtained by analysis of the percentage of the cases in which the difference in performance between the pairs exceeded zero. On Trial 9 of cancellation, for the Maximum level of aspiration, 54 percent of the mean differences exceeded zero, showing a slight tendency for more *S*s in the High aspiration group to perform Trial 9 more slowly than did their matches in the Low aspiration group. The P-value of the χ^2 for this percentage, however, is between .70 and .50, indicating that the tendency is unreliable. What is true of cancellation is also true of arithmetic, *i.e.*, consideration of the data on the Maximum level of aspiration for the three critical trials reveals that although the *S*s were reliably divided in terms of aspiration, there was little dependence of performance upon difference in Maximum level of aspiration. However, although none of the differences were reliable, there did seem to exist a slight tendency for high Maximum level of aspiration to be associated with more efficient performance in the ego-involved task.

In the analysis of the dependence of performance upon the Actual level of aspiration, there were 49 [3] pairs of *S*s matched for performance on Trials 6, 7, and 8, and for estimate of performance on Trial 8 (Table 3). Table 4 shows that for cancellation there was no significant difference in performance at Trial 9 although there was a reliable difference between the pairs in terms of aspiration. The trend was for those in the High group to produce a slower performance, but the P-value of the mean difference of 1.53 sec. fell between .17 and .16, and the tendency was, therefore, unreliable. No significant difference in performance existed for Trials 10 and 11. In the arithmetic task we found no reliable differences in performance, but the absolute

[3] The number of pairs differs from estimate to estimate because of the elimination of those matches who gave identical levels of aspiration.

values were somewhat more pronounced in the direction of faster performance being dependent upon higher level of aspiration. The mean differences in performance were −3.82 sec., −5.54 sec., and −4.42 sec. for the three critical trials, and the corresponding P-values were .14, between .07 and .06, and between .09 and .08. The differences in favor of the High aspiration group represent an improvement of approximately 4 percent over the Low aspiration group.

Fisher (**6**) has given the statistical procedure to be used in the situation in which, although independent tests do not yield a significant P-value, the aggregate of the individual P-values of the tests may be lower than would be expected by chance. The data on the dependence of performance upon the Actual level of aspiration in the arithmetic task lend themselves to this analysis, since the P-values at each trial, although not significant, are quite low, and in addition all differences are in the same direction. The sum of the values of χ^2 for each P-value in question is 14.0666 and its P-value (6 degrees of freedom) is between .05 and .02. It is apparent that, considering the data as a whole, the low individual P-values could have occurred less than five times out of 100 by chance.

When we consider the percentages exceeding zero in the case of the Actual estimates in arithmetic, we find that none of the measures are significant, although the direction was toward faster performance for those with high Actual aspirations on Trials 9 and 10. This particular measure is less satisfactory because it does not take into account either the absolute values of the performances or their variabilities.

In the case of the Least level of aspiration there were no significant differences in performance for the three critical trials in either cancellation or arithmetic. The trend, however, was for those with the higher Least levels of aspiration in cancellation to give poorer performances than their matches in the task. The mean differences in performance for Trials 9, 10, and 11 were .75 sec. (P between .35 and .33), .88 sec. (P = .09), and .39 sec. (P between .31 and .29). The percentages which exceeded zero for the three trials were 60 percent (P between .30 and .20), 57 percent (P between .50 and .30), and 56 percent (P between .50 and .30). In the arithmetic task the trend was for those in the High Least aspiration group to give the faster performances. The mean differences in performance were −1.28 sec. (P between .38 and .36), −.10 sec. (P between .42 and .40), and −.53 sec. (P = .44). The percentage exceeding zero was in the same direction only in the case of Trial 9, where it was 49 percent, with P coming between .90 and .80. For Trials 10 and 11, these percentages were 54 percent (P between .70 and .50), and 55 percent (P between .50 and .30).

The data contained in Table 4 lead to the conclusion that performance was not reliably dependent upon differences in Maximum or Least levels of aspiration in either task. There was no dependence of performance upon difference in the Actual level of aspiration in the non-ego-involved task. Although the P-values of the dependence of performance upon the Actual level of aspiration at the individual trials in the ego-involved task were not significant, there was a reliable indication that in terms of the aggregate of the trials higher Actual levels of aspiration tended to be followed by better performances. Apparently, in the case of individual *Ss* who differ in Actual level of aspiration in an ego-involved task no prediction can be made as to subsequent level of performance, but in *groups* of *Ss* we may expect a dependence between performance and Actual level of aspiration.

Dependence of Estimate of Past Performance Upon Levels of Aspiration.—In the analysis of the dependence of the estimates of past performance upon the levels of aspiration, the *Ss* were matched in the same manner as they were in the analysis of the dependence of performance upon the levels of aspiration. The data showing that the *Ss* were adequately matched are in Table 3. In the case of the Maximum level of aspiration for Trial 9 of cancellation, the 44 matched pairs showed a mean difference of .09 between the High and Low groups in estimate of performance on Trial 9 with a P-value coming between .44 and .42. For Trial 10 the mean difference was .33 (P between .15 and .14). No analysis could be made for Trial 11, since no estimates of past performance were obtained after the last trial in the experiment. It should be noted that on Trial 10 the mean difference was not reliable (P between .15 and .14), but 62 percent of the cases had differences which exceeded zero and this approaches significance, since P came between .05 and .02. For the arithmetic task there were no significant mean differences in estimate of past performance in relation to differences in the Maximum level of aspiration.

When we consider the estimates of performances 9 and 10 on the basis of differences in the corresponding Actual levels of aspiration we find reliable mean differences for both cancellation and arithmetic. In cancellation the mean differences were .72 and 1.00 for Trials 9 and 10, with the corresponding P-values being between .06 and .04, and less than .01. The P-value of the two trials taken as an aggregate is less than .01. In 73 percent of the cases the difference exceeded zero (P less than .01) on Trial 9, and 68 percent exceeded zero (P less than .01) on Trial 10. In arithmetic the mean differences for the two trials were 1.12 and .74, with P-values of less than .01, and between .02 and .01, respectively. On Trial 9, 71 percent

of the differences exceeded zero (P less than .01), and on Trial 10, 65 percent exceeded zero, with P falling between .02 and .01. The direction in every instance was for those who gave higher Actual levels of aspiration to follow them with higher estimates of past performance.

The estimate of past performance was not influenced by the preceding Least level of aspiration on either trial in cancellation. On Trial 9, arithmetic, the mean difference in estimate of past performance which followed the Least level of aspiration for that trial was .62 (P between .07 and .06), 66 percent of the cases had differences which exceeded zero (P between .05 and .02). It will be noted that here, again, we have a case in which reliability is demonstrated in one measure and not in another. However, on Trial 10 both measures were unreliable.

The data on the dependence of estimate of past performance upon preceding levels of aspiration warrant the conclusion that estimates of past performance are dependent upon preceding Actual levels of aspiration, regardless of task. Those individuals who gave the higher Actual levels of aspiration consistently tended to follow them with higher estimates of past performance.

Dependence of Estimate of Past Performance Upon Performance.—In the analysis of the dependence of estimates of past performance upon performance new sets of matchings had to be developed. In this instance it was necessary to match for the performance prior to the critical trial in order that a valid difference in critical performance would be produced. Since the estimate of past performance could be influenced by the previous estimate of past performance the latter had to be considered in the matching. Since it had been demonstrated that the estimate of past performance was influenced by the preceding Actual level of aspiration the *S*s had to be matched in terms of that measure. After matching for these three variables, the pairs were divided into High and Low performance groups. Since no estimate of past performance was obtained after Trial 11 the analysis considers only Trials 9 and 10. In cancellation for Trial 9 there were 47 pairs matched for the variables outlined above. The P value of the mean of the differences for previous performance (Trials 6, 7, and 8) was between .35 and .33 and the P value of the difference in variability was between .40 and .38. In the case of the estimate of what performance at Trial 8 had been the P value of the mean difference fell between .13 and .12 and the P value of the difference in variability between .27 and .26. For the Actual level of aspiration for Trial 9 the P value of the mean difference fell between .48 and .46 and the P value of the difference in variability was .50. It is apparent that the *S*s were matched on the three variables both in

terms of central tendency and variability. The P value of less than .01 for the mean difference in performance at Trial 9 showed that the *Ss* were reliably divided in terms of that performance. The mean difference of the estimate of what performance at Trial 9 had been was −.11 with the P value between .40 and .38, and 49 per cent of the cases gave a difference which exceeded zero, the P value being between .90 and .80. These data force the conclusion that the estimate of past performance in this case was not dependent upon the performance itself. Consideration of the data for Trial 10 cancellation and Trials 9 and 10 arithmetic shows that on none of these trials was the estimate of past performance influenced by the performance being estimated.

Dependence of Estimate of Past Performance Upon Previous Estimate of Past Performance.—The variables, other than the previous estimate of past performance, which immediately preceded the estimate of past performance at a given trial were the performance and the levels of aspiration for that performance. It has been shown that of these latter variables the only one which exerted an influence on estimate of past performance was the preceding Actual level of aspiration. Therefore, in studying the dependence of the estimate of past performance upon the preceding estimate of past performance it was necessary only to match for the Actual level of aspiration. In cancellation there were 31 pairs matched for the Actual level of aspiration for Trial 9. The P value of the mean of the difference between these pairs was .50 and the P value of the difference in variability was between .40 and .38, showing that any difference between the pairs was due to chance. The two groups were reliably divided in terms of the estimates of what performance on Trial 8 had been since the P value of the mean difference of that estimate was less than .01. The mean difference of the estimates of performance on Trial 9 was .81 with a P value between .04 and .03, and 66 per cent of the cases gave a difference which exceeded zero, but the χ^2 of this percentage was not quite reliable, the P value falling between .10 and .05. However, for Trial 10 cancellation and Trials 9 and 10 arithmetic both critical measures were reliable, all of the P values being less than .01. It is evident that the estimate of what performance was on a particular trial was dependent upon the estimate of what performance had been on the previous trial. The direction of this dependence was for high estimate of past performance to be associated with high previous estimate of past performance.

Discussion

The theory was advanced that the needs to do well and to avoid failure should operate in such a manner that the individual with the

higher Actual level of aspiration should have a greater incentive to perform more efficiently than one with a lower Actual level of aspiration, particularly in an ego-involved task. The results of this experiment justify such a conclusion only so far as group results are concerned.[4]

We offer the following explanations of the fact that the dependence of performance upon the preceding Actual level of aspiration was so small in amount. It is possible, first, that the arithmetic task did not create enough ego-involvement to produce a striking effect. It is possible, secondly, that these explicit statements of aspiration are not reliable measures of subjective goal-striving. This view is in keeping with opinions previously cited. We must admit that the differentiation of three levels of aspiration might still not be refined enough for the solution of our present problem. For instance, following Gould (12), the three levels can still be either expressions of desire to do well or of merely wanting to appear ambitious. Perhaps, as Irwin suggests, a key to future experimentation would be the shift from asking for *expectations* concerning future performance to asking the *S*s to state what they are *trying* to achieve. In the final analysis, from a practical point of view, future experimentation should acknowledge both alternatives.

Turning now to a consideration of all of the experimental results which bear upon the relations between aspirations, performance, and estimates of past performance we shall attempt to show how they may be used to develop a general theory of the behavior of the variables involved. We shall consider first the extent to which the Actual level of aspiration and the estimate of past performance reflect the operation of the same forces.

That some difference exists between these two variables is evident in the fact that the estimates of past performance are less variable than the Actual levels of aspiration.[5] This suggests that the estimate of past performance is more in the nature of a judgment than is the Actual level of aspiration.[6] The estimate of past performance, as an estimate of an event which has already occurred, is presumably based to some degree upon objective information already received. The use of fictitious scores results in all *S*s' estimating from the same

[4] Although Fryer (8) has made a distinction between interests and motivation, it would seem that our results are to some extent comparable to those obtained in experiments dealing with the relationship between interest and achievement. The results in this field have failed to reveal any high correlations between interests and ability (8). We are comparing here the two approaches because they are both based upon statements which are subjectively determined.

[5] The data comparing the variabilities and distributions of the Actual levels of aspiration and the estimates of past performance are in the author's doctoral dissertation "Interrelations between Levels of Aspiration, Performance, and Estimates of Past Performance," on file in the Psychological Laboratory, University of Pennsylvania.

[6] McGehee (21) has indicated that although the level of aspiration is to some extent a judgment it is more a product of the 'ego-level' than of the process of judging.

objective background, a circumstance creating homogeneity of estimation of past performance. On the other hand, the Actual level of aspiration is an estimate of an event in the future, and therefore has more freedom of determination.

Although the Actual level of aspiration and the estimate of past performance have been shown to be dissimilar in the above respects, there are many similarities between the two measures which suggest that to a considerable extent they are influenced by the same forces. The most striking fact relating the Actual level of aspiration and the estimate of past performance is the fact that the latter is heavily dependent upon the Actual estimate regardless of task. Apparently individuals tend to feel that their performances have come within the region of their immediate aspirations in both ego-involved and non-ego-involved tasks. The need to do well and to avoid failure, which is concentrated in the Actual level of aspiration seems to extend into the estimate of past performance, which can be viewed as an expression of a 'satisfaction-seeking' need in which one hopes he has *done* well and has *avoided* failure.

Further evidence in support of the position that the Actual level of aspiration and the estimate of past performance are the products of similar needs is seen in other striking similarities in their behavior. The variabilities of the two measures, although different, are of the same order of magnitude, the sigmas ranging from 2.38 to 3.34; on the other hand, the sigmas of the Maximum estimates range from 13.71 to 18.41, and those for the Least estimates from 13.82 to 32.85. Finally, the distributions of the Actual estimates and the estimates of past performance are both approximately symmetrical whereas the distributions of the Maximum and Least estimates appear markedly skewed.

In addition to the fact that the needs which influence the Actual level of aspiration and the estimate of past performance are highly similar the evidence suggests that these needs tend to persist in time. The reliability coefficients reported by Frank (**7**) and Preston and Bayton (**24**) indicate that there is a within-task tendency to place the levels of aspiration at characteristic heights. In other words, there are individuals who give characteristically high aspirations and others characterized by low aspirations. The results of the present experiment show that the same phenomenon exists in respect of the estimates of past performance. When the Actual level was held constant, and with performance *per se* playing no part, those subjects with the higher estimates of past performance on one trial maintained higher estimates of past performance on the next trial.

The foregoing discussion appears to justify the following theory of the nature of the interrelations between levels of aspiration, per-

formance, and estimates of past performance. The individual feels a need to do well and a need to avoid failure. These needs find critical expression in the Actual level of aspiration. The Maximum level of aspiration also reflects these needs but is not of immediate incentive value because it is by definition removed in point of time and refers to ultimate aims, whereas the Actual level is characterized by immediacy.[7] The Least level defines complete failure and is an avoidance goal, but success-feelings would not be gained merely by performing above the Least estimate, but would rather be gained in achieving one's immediate goal, namely, the Actual level of aspiration. The needs which are reflected by the Actual level of aspiration are not momentary but continue to exert their influence upon subsequent events, including the formation of new levels of aspiration and new estimates of past performance. In ego-involved tasks there tends to be a demand on the individual to satisfy these needs by the release of energy commensurate with that necessary to attain the goal expressed in the Actual level of aspiration. After the performance, regardless of the degree of ego-involvement and its determination of the release of energy in the performance, these needs are reflected in the attitude one has toward his performance which find expression in the estimate of past performance. Here the individual hopes that he has done well and avoided failure.

It is not uncommon to find individuals evaluated in terms of their expressed levels of aspiration. Personnel workers may be influenced by the aims stated by job applicants. We seem now to have objective data which bear upon the significance of statements of this kind. For one thing, the interpretation of the real incentive value of the aspiration must be made in terms of the meaning of the task to the individual. Apparently we can expect a relation between aspiration and achievement only in those tasks in which the person's ego is deeply involved. However, even in an ego-involved task it is our *immediate aspirations* which are of incentive value. It would seem, then, that the most important subjective areas which should be explored in this appraisal should be those having to do with the immediate levels of aspiration.

In conclusion I would like to point out that among psychologists there has been a growing feeling of dissatisfaction with static, questionnaire approaches to the study of personality (**30**). Recent years have been characterized both by demands, and attempts to satisfy these demands, for the analysis of personality manifestations through the use of functional and rigidly experimental methods. No one has denied that Hoppe introduced an important problem in personality

[7] Thomsen (**28**) has advanced the view that those individuals who expect maximum achievements in the immediate future are particularly courting unhappiness.

and recognized most of the major factors which influence our self-motivated behavior. Subsequent experimentation has further refined both the methods and the concepts involved until it seems that today substantial progress is being made toward the achievement of experimental methods for the study of personality.

Summary

The level of aspiration has been interpreted as being an expression of certain needs: the need to do well (or to appear to be wanting to do well); the need to approximate the future performance; and the need to avoid failure. Research up to the present time has been mainly concerned with the analysis of factors which influence the operation of these needs in the determination of the level of aspiration. The present experiment attempted to investigate the nature of the influence of these needs upon the events following the statement of the aspiration, namely, the performance for which the aspiration was given, and the individual's estimate of that performance.

There are at least three levels of aspiration available to the individual. The Maximum level of aspiration represents ultimate ability, the Least level of aspiration is the score below which the individual is certain he will not fall, and the Actual level of aspiration is the score which the subject expects to make on the next trial. The hypothesis was presented that in an ego-involved task the needs expressed in the Actual level of aspiration would continue to exert their influence in such a manner as to produce a dependence of height of performance upon the height of the Actual level of aspiration for that performance. A further hypothesis was presented to the effect that satisfaction-seeking arising as a consequence of the needs would operate in such a manner as to make the height of the estimate of past performance dependent upon the height of the Actual level of aspiration for that performance.

The present experiment was designed so as to permit a trial-by-trial test of the two hypotheses in an ego-involved task and a non-ego-involved task. The results support the view that the influences which find expression in the level of aspiration continue to be operative upon the events which follow the statement of the aspiration. Although the data were unreliable for individual trials, the data for the aggregate of the three critical trials, in the ego-involved task, revealed a reliable tendency for those subjects with higher Actual levels of aspiration to follow them with better performances. Also, regardless of the nature of the task, there was a reliable tendency for those subjects who gave higher Actual levels of aspiration to follow them with higher estimates of past performance.

(Manuscript received April 19, 1943)

Bibliography

1. Anderson, H. H., & Brandt, H. F. Study of motivation involving self-announced goals of fifth grade children and the concept of level of aspiration. *J. soc. Psychol.*, 1939, **10**, 209–232.
2. Bayton, J. A. The level of aspiration technique as an experimental method for the study of morale. *Virginia J. Sci.* (Abstract, in press.)
3. Chapman, D. W., & Volkmann, J. A social determinant of the level of aspiration. *J. abn. soc. Psychol.*, 1939, **34**, 225–238.
4. Festinger, L. Wish, expectation, and group standards as factors influencing levels of aspiration. *J. abn. soc. Psychol.*, 1942, **37**, 184–200.
5. Filter, R. O. Estimates of the amount of work one can do. *J. appl. Psychol.*, 1927, **11**, 58–67.
6. Fisher, R. A. *Statistical methods for research workers.* London: Oliver & Boyd, 1936.
7. Frank, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–128.
8. Fryer, D. A. *The measurement of interests.* New York: Henry Holt, 1931.
9. Gardner, J. W. The use of the term "level of aspiration." *Psychol. Rev.*, 1940, **47**, 59–68.
10. Garrett, H. E., & Schneck, M. R. *Psychological tests, methods, and results.* New York: Harper & Brothers, 1933.
11. Gould, R. An experimental analysis of "level of aspiration." *Genet. Psychol. Monogr.*, 1939, **21**, 3–115.
12. Gould, R. Some sociological determinants of goal strivings. *J. soc. Psychol.*, 1941, **13**, 461–473.
13. Gould, R., & Lewis, H. B. An experimental investigation of changes in the meaning of level of aspiration. *J. exp. Psychol.*, 1940, **27**, 422–438.
14. Hertzman, M., & Festinger, L. Shifts in explicit goals in a level of aspiration experiment. *J. exp. Psychol.*, 1940, **27**, 439–452.
15. Hilgard, E. R., & Sait, E. M. Estimates of past and future performance as measures of aspiration. *Amer. J. Psychol.*, 1941, **54**, 102–108.
16. Hilgard, E. R., Sait, E. M., & Magaret, G. A. Level of aspiration as affected by relative standing in an experimental social group. *J. exp. Psychol.*, 1940, **31**, 411–421.
17. Hoppe, F. Erfolg und Misserfolg. *Psychol. Forsch.*, 1930, **14**, 1–62.
18. Irwin, F. W., & Mintzer, M. G. Effect of differences in instructions and motivation upon measures of the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 400–406.
19. Kneeland, N. Self-estimates of improvement in repeated tasks. *Arch. Psychol.*, 1934, **163**, 75.
20. MacIntosh, A. Differential effect of the status of the competing group upon levels of aspiration. *Amer. J. Psychol.*, 1942, **55**, 546–554.
21. McGehee, W. Judgment and level of aspiration. *J. gen. Psychol.*, 1940, **22**, 3–15.
22. Pennington, L. A. Shifts in aspiration level after success and failure in the college classroom. *J. gen. Psychol.*, 1940, **23**, 305–313.
23. Preston, M. G. Use of the coefficient of correlation in the study of the D-score for the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 442–446.
24. Preston, M. G., & Bayton, J. A. Differential effect of a social variable upon three levels of aspiration. *J. exp. Psychol.*, 1941, **29**, 351–369.
25. Preston, M. G., & Bayton, J. A. Correlations between levels of aspiration. *J. Psychol.*, 1942, **13**, 369–373.
26. Rotter, J. B. Level of aspiration as a method of studying personality. *Psychol. Rev.*, 1942, **49**, 463–474.
27. Sears, P. S. Levels of aspiration in academically successful and unsuccessful children. *J. abn. soc. Psychol.*, 1940, **35**, 498–536.
28. Thomsen, A. Expectation in relation to achievement and happiness. *J. abn. soc. Psychol.*, 1943, **38**, 58–73.
29. Tiffin, J., Knight, F. B., & Josey, C. C. *The psychology of normal people.* Boston: D. C. Heath, 1940.
30. Woodworth, R. S. The adolescence of American psychology. *Psychol. Rev.*, 1943, **50**, 10–32.
31. Yacorzynski, G. K. Degree of effort: III. Relationship to the level of aspiration. *J. exp. Psychol.*, 1942, **30**, 407–413.

LATENCY OF RESPONSE IN A CHOICE DISCRIMINATION

BY HAROLD SCHLOSBERG AND RICHARD L. SOLOMON

Brown University

Introduction

Latency of response, or the interval elapsing between a stimulus and a reaction, has been widely used as a measure of the strength of a response or of a 'response tendency.' In the traditional reflexes of the physiologist, latency reflects stimulus strength, complexity of the neural pathways involved, and the state of these pathways. In the field of classical conditioning, latency has been used chiefly as an index of the strength of the conditioned response, making it possible to study the development of conditioning and extinction. In spite of Skinner's (**15**) early dissatisfaction with this measure, latency has also been used in order to follow the progress of so-called operant conditioning. Graham and Gagné (**3**) found that the logarithm of latent time furnished a satisfactory measure of reflex strength in their runway situation. Verplanck (**21**) extended the method to a black-white discrimination with the same apparatus. He used the method of single stimuli, never presenting the black and white runways at the same time. Verplanck was able to demonstrate most of the phenomena which Pavlov (**13**) found in his classical work with the salivary response. Finally, Solomon (**16**) was able to duplicate Verplanck's results in a clear-cut fashion, using latency of response in a single-door jumping situation. Perin (**14**) and Williams (**23**) have also used this measure.

The present study was concerned with the significance of latency in a simultaneous black-white discrimination. It was actually performed before Solomon's work with the single door, successive discrimination, and was designed to test a theory of discrimination developed by Schlosberg and Finger (**1**). Finger (**1, 2**) had used latency as one of the measures of response strength during a study of conflict in the white rat. In this study Finger used the traditional Lashley jumping situation, training the animals in a black-white discrimination. He produced conflict by substituting various gray cards for the training pair, thus making the discrimination more difficult.

In planning Finger's experiment it had been tacitly assumed that the conflict would be between two inconsistent courses of action, namely, jumping to the lighter card versus jumping to the darker

one. Most rats developed a type of behavior that is best described as VTE. (19). The rats would crouch on the stand directly in front of one card, often leaning forward as if to jump, and then either jump or move along the stand and repeat this performance opposite the other card. This alternate 'sampling' was repeated until the rat finally jumped. The separation of the doors and the behavior of the rat were such that the rat was obviously responding to only one card at a time. It seemed gratuitous to assume that he was making a comparison between the stimulation of one card and the 'image' or trace of the other, or that the actual conflict was between two different forward-going tendencies. It is much more probable that the response of the rat was determined by the single card in front of him *at the time* he crouched for the jump. The development of this view resulted in a theory of discrimination in which latency plays an important part. This theory involves four simple assumptions, all subject to eventual testing. For clarity the assumptions are presented in terms of the present experiment, which used a Lashley jumping stand, and five stimulus cards equally spaced along a white-black continuum.

(1) The positive training card (white) has acquired a strong positive excitatory value through its association with food during the training period.

(2) The positive excitatory value decreases along a generalization gradient as we go along the white-gray-black continuum.

(3) There is a negative factor which prevents the excitatory value of the stimulus from resulting in a response as soon as the rat looks at a given stimulus card. The negative factor may be simply the arresting effect of the 16-inch gap between the stand and reward. As long as no punished errors occur, the value of the negative factor will be determined by the distance to be jumped and will therefore be equal for all stimulus cards.

(4) While the rat remains on the jumping stand, the balance of positive and negative factors shifts toward the excitatory side, presumably as a logarithmic function of time. As soon as the positive excitatory value of either of the stimulus cards exceeds the negative factor, the rat jumps. The mechanism of the shift of balance may be thought of as simple temporal summation of the excitatory value of the stimulus card, but additional work is necessary to clarify this point.

These four assumptions give a reasonable explanation of the relationship of latent time to strength of response tendency in the single-door jumping situation. Indeed, some such assumption seems necessary if we are to do anything beyond merely defining latency as a measure of response strength. The gradients referred to under

(2) above have been measured in the single-door jumping situation, and present a surprisingly consistent picture (16).

The four assumptions we have made will also explain choice discrimination, in which two stimuli are present, as in the experiment to be reported here. The only difference is that there are two stimulus cards, differing in excitatory potential. When the rat is placed on the jumping stand neither stimulus arouses enough excitatory potential to overcome the strength of the negative factor, so that the rat does not jump immediately. But while the rat remains on the stand, the balance between excitatory potential and the negative factor gradually shifts, and as soon as the excitatory value of one of the stimulus cards exceeds the negative factor the rat will jump to that card. The rat will then be expected to jump to that card which has the highest excitatory value on the generalization gradient, not by comparing the cards directly, but because the latent period of a jump to that card is shorter than the latent period of a jump to the other card.

This oversimplified account undoubtedly needs further additions and qualifications which we shall discuss later. But it offers several clear-cut predictions which can be tested. Many of the predictions are quite obvious,[1] concerning the effect upon latent period and on sharpness of discrimination of such factors as strength of drive, length of gap to be jumped, etc. The most striking prediction is the one to be tested in the present study, and may be stated as follows: In a stable choice discrimination, the latency of response will depend upon the positive stimulus, and will be independent of the negative stimulus. The prediction is interesting because it runs directly counter to common sense, which generally assumes that the closer two stimuli are to each other along a continuum the slower will be the choice reaction time.

Apparatus, Stimuli, and Procedure

Apparatus

Finger's adaptation (1) of the Lashley jumping stand was used for our experiments. Rats were taught to jump from a stand to a reward platform, through one of two openings in a vertical panel attached at right angles to the front of the platform. The openings could be covered with stimulus cards. The cards were hung at the top of the square openings, so that they swung back easily when the rats jumped to them. The incorrect stimulus could be backed up by locking a door behind it. If the rat hit this, he fell into a cloth apron below. Since Finger has presented detailed dimensions of the apparatus, it seems unnecessary to do so here.

In a preliminary experiment it was found that rats often jumped from a point midway between the two openings, and hit the inner edges of the doors. A 6 in. x 6 in. piece of 1/2 in. wire mesh was placed upright at the center of the front edge of the jumping stand, to force the rats to jump from points directly in front of the openings. In place of a single central food box,

[1] In the sense that any theory would predict them. Until we have more quantitative data on generalization gradients, these predictions cannot be made precise enough to serve as crucial experiments.

a separate one was placed opposite each opening. These two changes resulted in jumps through the centers of the windows, remedying the original difficulties.

Stimuli

The stimulus cards were 6 in. x 6 in. pieces of paper, mounted on cardboard. Educo Mimeograph paper and Hering Gray No. 29 were arbitrarily chosen as white and black. Three intermediate steps of gray were chosen by a fractionation method. White and black were placed 12 in. apart on a black background, and each of five graduate students was asked to find that Hering gray which seemed halfway between them in brightness. The subjects were allowed to try several samples in the space between white and black before making a final choice. All picked No. 13 as middle gray (M.G.). When the black—middle-gray interval was halved by the same procedure, four subjects chose No. 20 and one chose No. 22. Thus No. 20 was called dark gray (D.G.), although it may be slightly too light. Similarly, four students agreed on number 6 as light gray (L.G.), while one chose No. 4, suggesting that our light gray might be slightly too dark. The relative brightnesses of the five cards were determined with a MacBeth Illuminometer, and are given in Table I. The cards are quite evenly spaced along the white-black continuum in terms of the logarithms of relative brightnesses. Thus the cards yield equal intervals to the human eye, and give equal differences in log reflectance values, though they are not quite equally spaced in terms of the Hering numbers.[2]

TABLE I

MacBeth Illuminometer Readings *

Stimulus Card	Hering No.	Brightness in Ml.	Relative Brightness	Log Rel. Brightness
White		.147	100%	2.00
Light Gray	6	.078	53%	1.72
Middle Gray	13	.035	24%	1.38
Dark Gray	20	.021	14%	1.15
Black	29	.012	7%	.84
Background		(.002)	1.4%	.15

* With the exception of the background the absolute brightnesses given in the table were obtained under the actual conditions of illumination used in the experiment. Relative brightnesses were also checked at higher intensities and the value for the background was calculated from these readings.

Experiment I

Establishing the Discrimination

Four male litter-mates were used. They had been reared in the laboratory colony (inbred, chiefly of Wistar stock, with no pigmentation) and had been "gentled" by frequent handling throughout their lives. They were 90 days old when experimentation started. Six days were spent in adjusting the animals to one unlimited feeding each evening. Then three evenings were devoted to training the rats to walk through the doors and eat in the foodboxes. Only one door, chosen at random, was open on each trial, and the jumping stand was placed in contact with the front panel so that the rats could walk through to food with no difficulty. The reward was 1/4 gm. of wet Purina dog chow, fed to the rat in either food box. Thirty trials were generally given in one evening. Thus the total amount of mash eaten during a session was about 7.5 gms., which was not enough to cut down the hunger drive appreciably. After each session the rats were allowed to eat to satiety in a food box.

[2] Unfortunately, in the actual experiment, we depended upon the black paint which covered the doors for our negative stimulus. This looked about as black as Hering No. 29, but had a more glossy texture. Actual Illuminometer readings showed it to have considerably less reflectance than No. 29, which would put it about two fractionation steps darker (by extrapolation). The matter is probably of no great importance, since black was always the negative stimulus, and was thus a constant.

At the end of three such sessions the rats were no longer exploring the apparatus, but were running promptly from the stand through the open door to food. The stand was then moved progressively away from the panel, until the rat was jumping 16 inches to the doorway. Then a piece of white paper was hung behind the doorway, so that it covered the upper half of the opening, but would swing out of the way when the rat hit it. As the rat learned to jump through the paper it was lowered until it covered the entire opening. At this point the rat was jumping in a white-black discrimination with 100 percent accuracy, with no history of errors. Ten sessions in all had been devoted to this preliminary training. White cardboard was then substituted for the paper, and 19 sessions were devoted to stabilizing the discrimination.

Throughout this period latencies had been recorded with a silent stopwatch (1/5 sec.). The watch was started when the rat was placed on the stand, and stopped when he jumped. Latencies had fallen rapidly during the first five of the white-cardboard sessions, approaching a value of about 3–4 sec. There was no further decrease during the last 14 preliminary sessions, so the black-white discrimination may be considered well stabilized. It should be emphasized that the training was gradual, so that the rats never made an error. Thus the gradient established by reward was uncomplicated by the effects of "punishment."

Test Series and Results

Each of the four animals was subjected to 12 successive test nights, of 32 jumps per session. The following pairs of cards were tested: light-gray—black, middle-gray—black,dark-gray—black, white—dark-gray, and white—middle-gray. Each of these five pairs was tested each night, interspersed among 27 presentations of the training pair (black-white). Some effort was made to balance out order and position effects, but the small number of animals made it impossible to do this adequately. It seemed more important to limit the number of test jumps in each session, always separating them by five training pairs, to minimize the effects of test jumps on generalization.

Latencies.—The latencies for each pair during the 12 test sessions are presented in Table II. In the upper portion of the table we

TABLE II

SUMMARY OF RESULTS

Discrimination	Mean Log Latency		Geometric Mean	
	Exper. I	Exper. II	Exper. I	Exper. II
Test Jumps, First Night				
W—B	.67	.527	4.68 secs.	3.36 secs.
LG—B	.94	.657	8.71	4.54
MG—B	1.27	1.370	18.60	23.40
DG—B	1.31	....	20.40	....
W—DG	.79	.667	6.21	4.64
W—MG	.79+	.737	6.17	5.46
LG—DG	...	.687	....	4.86
Test Jumps, All Nights				
W—B	.64	.536	4.33	3.44
LG—B	.80	.631	6.28	4.28
MG—B	1.03	1.189	10.70	15.40
DG—B	1.20	....	15.70	....
W—DG	.71	.617	5.11	4.14
W—MG	.74	.637	5.48	4.33
LG—DG	...	.633	....	4.30

have presented the average results for the first session, since this furnishes measures of initial generalization. (But see Discussion, below.) In the lower portion of the table are the averages for all 12 sessions. These values presumably reflect some influence of repeated testing. All values are presented in terms of mean log latencies, and geometric mean latencies in seconds (anti-log). The use of mean log latencies had been defended by Graham and Gagné (3), Finger (1) and Solomon (16). The geometric means are presented for convenience. An examination of the table will show that all pairs containing white as the positive card yield shorter latencies than those pairs in which one of the grays is positive to black. Thus the latency for W—MG on the first night is shorter than that for LG —B, although the former interval represents two fractionation steps on the continuum while the latter represents three (or more: see footnote 2). This statement holds for the averages of the 12 nights as well as for the first night.

VTE.—Vicarious Trial and Error Behavior (VTE) was recorded during the 12 test sessions. An arbitrary score of one VTE was counted when the rat started toward stimulus A, moved toward B, then returned to A. VTE scores for all rats during 12 test sessions were distributed among the various discrimination pairs, and expressed in terms of percentage of occurrence. (See Table III.)

TABLE III

VICARIOUS TRIAL AND ERROR

Stimuli	Number of Occurrences		Relative Occurrences	
	Exper. I	Exper. II	Exper. I	Exper. II
W—B	2	2	1.4%	2.6%
W—MG	15	0	10.6%	0 %
W—DG	5	0	3.5%	0 %
LG—B	2	3	1.4%	3.8%
MG—B	38	48	26.9%	61.5%
DG—B	79	..	56.0%	.....
LG—DG	..	25	.....	32.0%

Thus, for example, of 141 observed occurrences of VTE, two, or 1.4 percent, took place on the W—B pair. In general, the incidence of VTE parallels the latency measures, except that VTE values for W—MG seem too large, while those for LG—B seem very small. The amount of VTE is, of course, *too small to make such comparisons more than suggestive.*

EXPERIMENT II

Four rats were brought to a stable level of black-white discrimination as in Experiment I. Twelve additional sessions were then devoted to shifting the discrimination to one between

light gray and dark gray. This was done gradually, by first substituting Hering Gray No. 28 in place of black, and No. 2 in place of white. When this was mastered the pair was changed to No. 27 versus No. 3, and so on, until the rats had reached a stable level of latencies on light gray (No. 6) versus dark gray (No. 20). One rat was discarded during this procedure because he made several inaccurate jumps which threatened to complicate the results. The remaining three animals were then subjected to test jumps. The tests differed from those of Experiment I in the following ways: (1) LG—DG was substituted for DG—B; (2) each pair was presented twice in each session, once with the positive card on the right, and once with it on the left; (3) four jumps on the training pair (instead of five) came between each test pair; and (4) the testing program was extended to 16 sessions.

Results

Latencies.—Latency values are presented in Table II. It will be noted that all latencies were decreased, except for MG—B which was increased. The interpolated training had made the members of the LG—DG pair more nearly equivalent to W—B, respectively.

VTE.—See Table III. MG—B yielded 62 percent of all VTE. LG—DG, one of the training pairs, gave 32 percent. The remaining 6 percent were divided between W—B and LG—B. More occurrences on W—MG are noted, although the interval is presumably equal to MG—B, which yielded the greatest percentage of VTE. These results are similar to those of Experiment I.

Effect of position.

Analysis of variance showed no preference for left or right. This validated our use of one test per session on each pair, in Experiment I.

Discussion

In Fig. 1 we have plotted the mean log latencies of the four rats of Experiment I in a solid line, and of the three rats of Experiment II in a broken line. The baseline represents the fractionation values of the positive cards, while the ordinate represents mean log latencies when each card is paired with black. If we connect the points of either experiment we have a generalization gradient, presumably representing some aspect of the strength of the response tendencies associated with each of the positive cards. Values for the pairs in which the negative card is LG, MG, or DG do not contribute to the measurement of this gradient, and cannot be fitted to the baseline. They are plotted as points on the ordinate for convenient reference.

Let us first consider the results of Experiment I. A straight line furnishes a surprisingly good 'fit' to these points, especially when we consider that the baseline is made up of intervals which seem subjectively equal to the human observer. As a matter of fact, an equally good fit is obtained when the log latency values are plotted against log relative brightnesses of the cards as a baseline.[3] In

[3] It seems unnecessary to present this figure here.

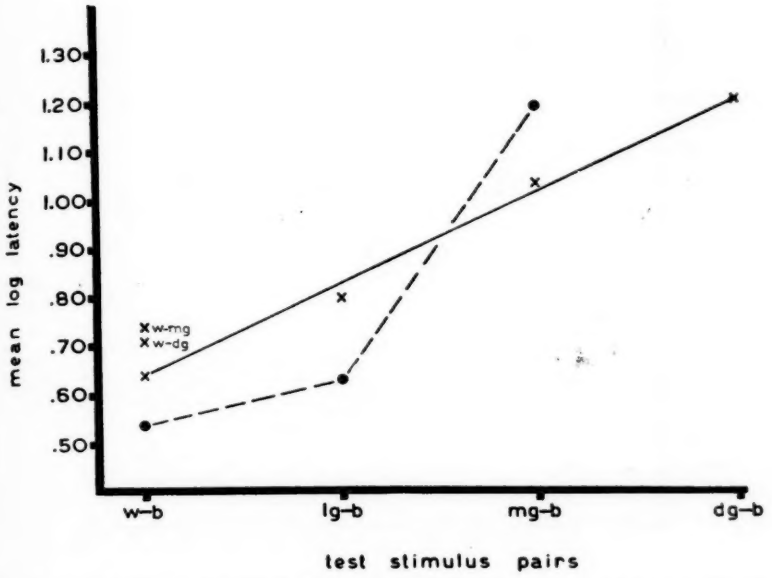

FIG. 1. The probable shape of the generalization gradient, in terms of mean log latency. The stimulus pairs (indicated on the base line) were chosen so that light gray, middle gray and dark gray mark off equal-appearing intervals along the white-black continuum. The solid line represents the probable form of the primary generalization gradient since the rats were trained on W—B (Experiment I). The two additional points above W—B represent values for white—dark-gray and white—middle-gray, respectively. The broken line shows the influence of additional training on LG—DG (Experiment II).

neither case do the points depart significantly from the best-fitting straight line. When we consider that there were only 48 measurements at each value (except white-black, for which there were 300) it seems reasonable to consider the 'true' gradient in Experiment I to be a straight line. This interpretation is strengthened by the sample daily gradients plotted in Fig. 3. These are derived from the 'learning curves' of Fig. 2. The gradient for each test session is plotted exactly as was the average for the 12 sessions in the previous figure. An examination of Fig. 3 will show that most of the daily gradients approach straight lines. It must be remembered that each plotted point represents the mean of four jumps, one from each animal for that night. The placement of the individual points is therefore quite unreliable. One cannot tell, then, whether the departures from linearity are significant or whether they are due to minor variations in the condition of the animal at the moment. The most probable departure is on the dark-gray—black pair on the first and possibly the second day (see below). But aside from this,

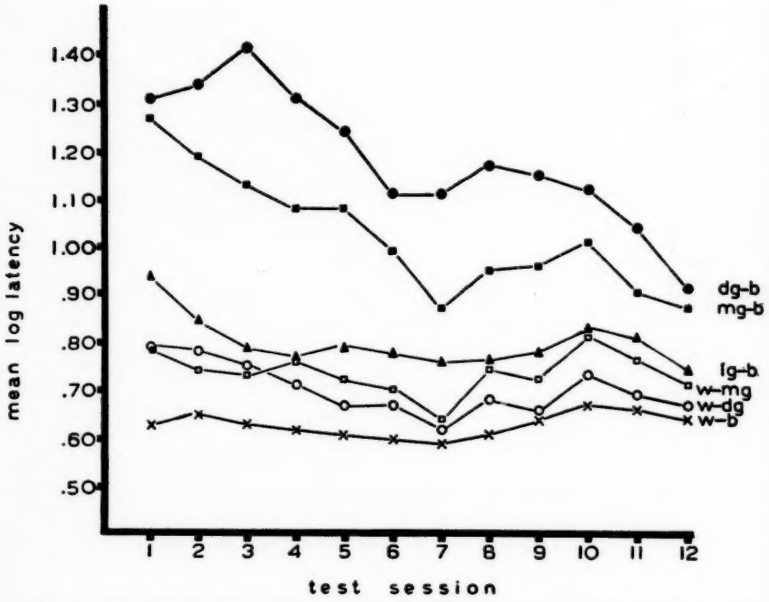

FIG. 2. Mean log latencies of jumps to the indicated stimulus cards on each of 12 successive test sessions of Experiment I. Each point (except W—B) represents the average of 4 test jumps. The points are connected to aid inspection.

the linearity of the daily gradients supports that obtained from averages of the whole 12 days.

There are at least two concepts that are frequently confused in discussions of generalization gradients. The first is the *primary generalization gradient*, which is probably determined by the receptors and their associated neural mechanism. The second concept is the *gradient of response strength*, which is due to the spread of the effects of punishment and reward along the primary generalization gradient.

The primary generalization gradient probably corresponds to what we call 'similarity,' and can be measured in man by psychophysical techniques, as the method of equal appearing intervals. Such intervals become appropriate steps for plotting gradients of response strength (**7**, **8**, **10**). The primary generalization gradient must be measured indirectly in animals. If we reinforce only one end of a continuum, we would expect the resulting gradient of response strengths to give a straight line, provided that response strength is measured in appropriate units,[4] and that the points tested

[4] We would define an appropriate unit of response strength as one that yields a symmetrical distribution of values around the mean at several points on the scale of strengths. This is our chief justification for using log latent times (*cf.* **2**, **3**). It would be interesting to see this test applied to Hilgard's (**6**) scale derived from an acquisition curve.

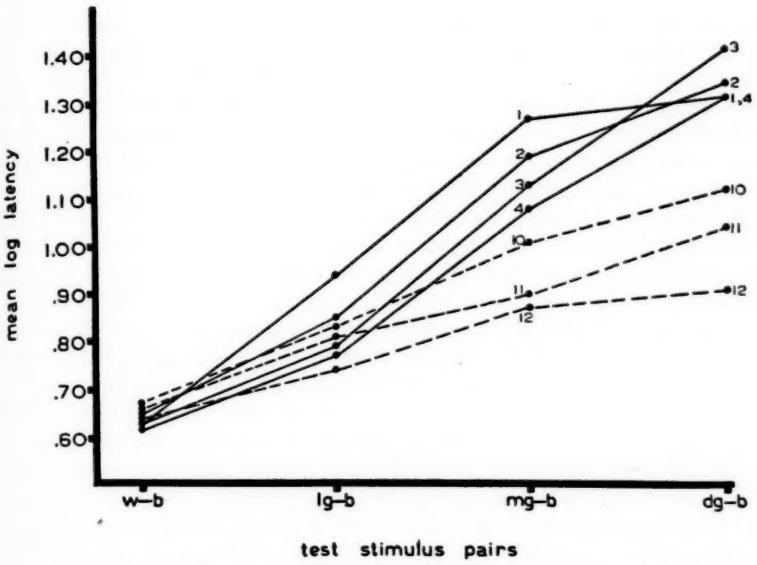

FIG. 3. Daily mean log latencies. Data from Fig. 2, days 1–4 and 10–12, are plotted on the ordinate and abscissa used in Fig. 1. Since most of the points are determined by only 4 jumps, the connecting lines are not accurate representations of daily gradients of generalization, but merely indicate trends and the cumulative effect of test jumps.

along the continuum constitute equal intervals. In other words, we would define equal steps along a continuum as those which yield a linear gradient of response strengths. This definition seems a bit circular. Its justification will be found in how consistently it relates various experimental data.

The present results offer some support for this line of reasoning. We reinforced the white end of a white-black continuum. We tested points along the continuum that fell at equal intervals in terms of their appearance to human subjects, who might be expected to have a mechanism of brightness discrimination similar to that of rats. The intervals were also essentially equal in terms of log relative reflectance, and of Hering Gray numbers. Response strengths, in terms of log latency, fell on a straight line, as we would have expected from our definition.

A more direct test of the usefulness of this definition of primary generalization gradient would be a plot of a gradient of response strength in man, since equal-appearing intervals can be measured directly by psychophysical methods. Hovland (**7, 8**) has carried out two such experiments. He determined the response gradient in terms of the amplitude of the psychogalvanic reflex for points

separated along the loudness continuum by 25 j.n.d.'s. The gradient is fairly flat but slightly concave. He points out that he would have used a scale of equal loudnesses if it had been available, but that any error introduced by the j.n.d. scale is cancelled by averaging response strengths obtained from two groups of subjects, one reinforced at each end of the continuum. But such averaging assumes equality of intervals, and is not legitimate. We have replotted his data for the two groups, determined equal distances in sones,[5] and then interpolated and averaged amplitude values for these equal distances. The resulting gradient is slightly convex, rather than concave.

Hovland carried out a similar study for pitch, and again obtained a flat, negatively-accelerated gradient of response strength. Here j.n.d. units furnish reasonably equal-appearing intervals, as he demonstrated experimentally, so there is no need of replotting. But Humphreys (**10**) criticizes j.n.d.'s as equal units on a pitch continuum on the basis of his own experiment, which showed octave generalization. He also presented gradients for shorter ranges of the pitch continuum. As a matter of fact, it is not certain that any of Hovland's or Humphreys' curves depart significantly from linearity. Nor can we be sure that equal increments in amplitude of the PGR correspond to equal changes in response strength (see footnote 4). Therefore none of these experiments gives us an unequivocal representation of the shape of the primary gradient. It should be relatively easy to design a series of experiments that would clear up the shape of the primary generalization gradient. The senior author hopes to set up a reaction time experiment analogous to the present one with rats. Another possibility is a refinement of Hovland's technique, with more attention devoted to the equality of units along the ordinate and abscissa. A series of grays, lines,[6] or lifted weights might serve as relatively simple continua. At any rate, the shape of the primary generalization gradient is probably important enough to warrant considerable effort to determine it. Now let us turn to a consideration of gradients of response strength.

[5] Calculated from Stevens and Davis (**18**), Fig. 43, p. 118.

[6] In 1906 Henmon, a student of Cattell, reported discrimination reaction times for graded series of hues, pitches and lines (**5**). The last of these series would seem to furnish a favorable continuum, for a simple scale of centimeters would correspond exactly to one constructed by equal-appearing intervals. Henmon had two subjects respond by pressing a key on the side of the longer of two lines. The standard was 10 mm. and the comparison lines were 10½, 11, 11½, 12, 12½ and 13 mm. We plotted Henmon's data in terms of the logarithms of mean reaction times (as in our Fig. 1) and obtained a slightly concave gradient. Unfortunately we could not calculate mean log latencies from the data, so we do not know how much the curve was distorted by probable skew of the distribution of raw reaction times. Further, the gradient is one of response strength, rather than a primary sensory one, since 'reinforcement' had been given repeatedly to all but one point (10 mm.) on the continuum.

In Experiment I the conditions were deliberately arranged so that the primary generalization gradient and the gradient of response strength would have the same form. But this parallelism need not always hold. For example, the DG—B latency on the first day (see Fig. 3) and probably on the second, is much too low. This may be a chance variation, or it may indicate the distraction by unfamiliar cards on the first night. (This point is elaborated by Solomon [16].) It is generally considered advisable to take the values obtained on the first night of test presentations as the measure of a gradient of generalization, since they do not reflect the effects of reinforcement of test stimuli. But we believe that the third or fourth night might give a more typical picture for the primary generalization gradient in this experiment. We used only one test on each of five cards each night, interspersed among 27 presentations of the training pair, and after some 600 jumps on the training pair. One would not expect these few tests jumps to affect the gradient very promptly. The actual results bear out this expectation, for the change in gradients from day to day is slight. But there still is a steady shift in the slope of the gradients, so that at the end of the 12 test sessions the latencies for generalized stimuli have dropped considerably.

A second factor which we found to influence the shape of the gradient of response strength is the type of original training. In Experiment II we gave 12 nights of training on LG—DG pair, after W—B discrimination had been stabilized. The average results are shown in the dotted line in Fig. 1. Unfortunately we did not introduce a DG—B discrimination pair because of the danger of errors on the part of the rats, so that the dark end of the gradient was not determined. But the three points tested show that the LG—DG training markedly increased the strength (decreased the latency) of the LG—DG pair. The gradient is almost level between W and LG, then rises steeply to MG. It probably would have flattened out again at the dark end if we had measured it. The essential point made clear by Experiment II is that there is no typical gradient for any continuum which is independent of reinforcement procedures. The gradient is established between the two training stimuli, even though they may not be at the extremes of a continuum (10). Thus the closer together the training pair the sharper will be the gradient between them, presumably indicating a better discrimination. In passing, it might be suggested that the use of latency opens the way to the measurement of various shapes of gradients created by specific training on the jumping stand. By appropriate selection of stimuli to be reinforced one might construct

a flat gradient, a steep gradient, concave or convex gradients, or even one with a bump in the middle of a continuum!

The Significance of Latency in a Choice Discrimination

During the last few years there has been considerable discussion of the theory of choice discrimination. In general there are two schools of thought. One is represented by Hull (**9**) and Spence (**17**), and has attempted to explain choice discrimination in terms of positive and negative gradients of generalization which are built up gradually during the training period. The theorists in this group depend heavily upon principles derived from studies of conditioning. The other school is not so easy to characterize, for its members do not always agree among themselves. Tolman has developed a rather elaborate description of discrimination involving such concepts as vectors, tensions, hypotheses and needs (**19**). Lashley (**12**), Krechevsky (**11**), and Haire (**4**) have supported different versions of a theory which assumes that several hypotheses or attempts at solution guide the animal's behavior during the acquisition of a discrimination. We cannot devote space here to the details of these theories, or to the arguments between their proponents, except that we wish to point out that all of them seem to assume that the animal directly compares the figures in a choice discrimination. This assumption is a natural one to make, since there are two or more stimuli present, and the animal responds to one of them.

But is such an assumption necessary? We must remember that there is a large literature on discrimination in a situation in which only one stimulus is presented at a time, and the animal 'discriminates' by responding or not responding. In the introduction to this paper we suggested that the mechanism of discrimination in the single stimulus and choice situation is the same.[7] We assumed that there is no direct competition between stimuli-released jumping tendencies, but that the real conflict is between a forward-going tendency released by the stimulus, and an opposing tendency not to jump. There are thus two conflicts, instead of one, in a two-stimulus situation.

In the single stimulus situation, as the rat stands before a figure, the effect of the positive figure summates, eventually resolving the conflict and producing a jump.[8] The length of time this resolution

[7] As long ago as 1912, Weidensall (**22**) pointed out that in a white-black discrimination against a black background, 'only the white is used' by the rat.

[8] Our first hunch was that negative tendency could be considered as a threshold effect. The threshold for jumping might decrease as a function of time elapsing since the last jump, analogous to stimulation during the relative refractory phase. But in an incidental experiment we found that inter-trial intervals of 30 to 60 sec. yielded identical latencies. It is thus *time on*

will take is the latent time, and will vary inversely as the number of reinforcements of the stimulus and its position on a gradient. It would also vary inversely with the strength of drive, and directly with the distance to be jumped and with any previous punishment. In the choice, or two-stimulus, situation we assume that the same thing happens, except that each figure will have its typical latency, and that the rat will jump to the more positive figure because the latent time has expired first.

On the basis of this theory we predicted that the positive figure would determine the latency in a given discrimination, and we designed this experiment to test the prediction. We have already discussed the shape of the generalization gradients, and we found that if the log latent time of jumps is plotted (jumps made to W, LG, MG, and DG, paired with B) we get a straight line (see Fig. 1). Now, if our theory agrees essentially with the results, latent time of W—DG and W—MG should be essentially the same as that of W—B, for all have the same positive card. On the other hand, the usual gradient theories would predict by inference that the latency would depend upon the difference of gradient values between the members of the stimulus pair. Thus one would expect W—DG to be equivalent in latency to LG—B, since each involved three fractionation steps, and presumably the same distances along our linear gradient. The experimental value for W—DG falls about halfway between the predictions made by the two theories. The value for W—MG is more favorable to our theory, for it is not clearly different from that of W—DG, and fairly close to that for W—B; but it is very far below MG—B, which has the same gradient difference between its members in terms of log latency. The same relations are maintained throughout the tests (Fig. 2). The results on the crucial pair would thus seem to be more favorable to our theory. But they also indicate that the theory is too simple and must take into account another factor. We believe that VTE is that factor.

the stand that must be considered. We have arbitrarily assumed that the tendency-not-to-jump remains constant during the latent period, and that the forward-going tendency increases through simple temporal summation of the effect of the viewed card. Other stimuli (*e.g.*, panel, jumping stand, and even negative card) undoubtedly contribute some positive effect which summates with that of the positive card, through the principle of sub-goal reinforcement. We are not able to measure these effects at present.

The negative tendency must not be discarded, however, for it is an important part of our theory. Its absence may explain those operant responses in which latency does not furnish an adequate measure of reflex strength, as Skinner's lever-pressing (15) and Keller's light avoidance (personal communication). The negative tendency is obvious in the Graham-Gagné runway, for rats hesitate to run out of a small dark box and proceed along a narrow runway in a lighted room.

Vicarious Trial and Error

When a trained rat is placed on the stand with a W—B pair in front of him there is little or no VTE; the black card receives little attention, and the rat goes almost immediately to the white one and crouches, preparing to jump. The lack of attention to the black is partly explained by its low value on the gradient of response, and partly by the fact that it does not 'stand out' from the uniform and equally black background furnished by the panel. Hence all the time spent on the stand contribute to the summating effect of the white stimulus. But if white is paired with a stimulus that has some positive excitatory value such as MG or DG, in a certain proportion of the trials the rat will explore the negative stimulus, eventually reject it, and then move to the white. But the time spent in front of the negative stimulus will add to the latent time of that jump, for it is still necessary for the animal to crouch in front of the white card for its normal latency period. This would explain the observed fact that latencies for W—DG and W—MG are slightly longer than those for W—B, without forcing us to abandon our theory that the latent time in a choice discrimination is determined by the length of time it takes the more positive stimulus to overcome the 'tendency-not-to-jump.'

VTE data from Experiment I bear out this assumption. The percentage occurrence of VTE for the three crucial discrimination pairs is: W—B, 1.4 percent; W—DG, 3.5 percent, and W—MG, 10.6 percent. Similar figures for Experiment II do not completely agree, for the percentages were: W—B, 2.6 percent; W—MG, 0 percent; and W—DG, 0 percent. But we must remember that our criterion for VTE was inadequate for tests involving white, since we defined a VTE unit as a double change; that is, going to card A, then to B, and then back to A again. The type of VTE we would expect on a pair containing white would involve going to or looking at the negative card, and then turning to the positive card and staying there. We would not expect the rat to leave the white card once he got there.

One cannot invoke VTE as an explanatory theory without explaining VTE. Tolman's 'schematic sowbug' (**19**) does not fit into our present system of describing choice discrimination, for the amount of VTE that the sowbug does depends upon the difficulty of discrimination, which, in turn, depends upon the difference in response strength existing between the two stimuli presented. We believe that it is equally possible to explain VTE without recourse to the assumption of direct comparisons on the part of the rat. In our description of choice discrimination we have essentially two

independent pairs of 'approach-avoidance' conflicts. If the 'approach' value for a given card is low, the animal will 'go out of the field' and move over to the other card. If the second card also has a low excitatory value, the animal will again go out of the field and return to the first card. This will continue until one card or the other exceeds the tendency not-to-jump. This sequence of events, as we have described it in conceptual terms, is, of course, VTE.

This theory of VTE is fairly consistent with our results. One half (56 percent) of the VTE in Experiment I occurred in the DG—B pair, in which neither stimulus has a very strong excitatory value (as determined from the generalization gradient). The next weakest pair, MG—B has 27 percent of all VTE. The results of Experiment II are equally marked. Tolman and Minium (**20**) found the same thing: DG—B yielded more VTE than did LG—B, with W—B evidencing the least. This was true for an overlearned discrimination, very similar to ours. Tolman had previously shown that in *initial learning* rats "do more VTE for an easy black-white discrimination than for a difficult white-gray discrimination" (**19**). White—medium-gray discrimination yielded an intermediate value. Our theory of VTE would not have predicted some of the results in Tolman's earlier experiment. But it does predict the large amount of VTE occurring early in learning, when the excitatory potential for the positive stimulus is weak. However, it does not explain the sudden increase in VTE, accuracy and latent time that occurred at about the eighth day of training. The data would lead us to expect a sudden generalized tendency not-to-jump; but we see nothing in the described conditions of the experiment to produce it. The actual gradients formed in Tolman's experiment may well have been complex, depending upon the brightness of the background, the preliminary training, and the range reinforced. Until these details and their effects are determined, we hesitate to discard our otherwise satisfactory theory of VTE.

Predictions

It is rapidly becoming the fashion to make predictions on the basis of one's favorite theory. We are not in complete sympathy with this trend. Rarely do the predictions set up the classical 'experimentum crucis,' for few predictions are precise enough to prove or disprove any reasonable theory. As factual material accumulates, psychological predictions may improve. But since predictions are more dramatic than new explanations of established phenomena, we are tempted to present these derivations from our theory:

I. Anything that increases the general level of forward-going tendencies (as, e.g., increased hunger) will decrease latencies, decrease VTE, and increase the number of errors.

II. Anything that increases the general tendency not-to-jump (as, *e.g.*, increased distance) will increase latencies, increase VTE, and decrease errors.

III. A higher level of discrimination can be attained if all the forward-going tendency comes from the stimuli to be discriminated, without the addition of urging by *vis a tergo* (such as airpuffs, tail-tapping). Such urging simply cancels some of the negative tendency, cutting down the latencies for all stimuli.

These three predictions should hold as long as no punished errors have occurred during the training. We do not know enough about generalization gradients from punishment to predict beyond this point.

Summary and Conclusions

I. Four rats were trained to a stable latency level for a black-white discrimination, using the Lashley jumping technique. Training was so gradual that no errors occurred during the experiment.

II. The rats were then given the following test pairs: W—B, LG—B, MG—B, DG—B, W—MG, and W—DG. These stimulus cards were spaced evenly along the white-black continuum in terms of equal-appearing intervals (human) and log relative brightnesses.

III. When plotted against the intervals, the logarithms of the latencies of response of the test jumps yielded a straight line. We believe this to be the typical primary generalization gradient, which acts as a basis for the development of a superimposed gradient of response strength.

IV. In a second experiment the preliminary training was completed on a LG—DG discrimination. Test jumps showed that this training reduced the latency of responses involving LG as a positive member, changing the shape of the gradient of response strength.

V. The results are explained in terms of the single-stimulus type of discrimination (operant) without recourse to the assumption of comparison between stimuli. That is, the rat jumps to the first stimulus whose excitatory value exceeds the jump threshold at the moment.

VI. VTE roughly parallels latency of response. It is suggested that VTE occurs when neither stimulus has an excitatory value that is high in relation to the tendency not-to-jump.

(Manuscript received March 29, 1943)

Bibliography

1. Finger, F. W. A quantitative study of the effects of visually induced "conflict" in the white rat. Unpublished Ph.D. Dissertation, John Hay Library, Brown University.
2. Finger, F. W. Quantitative studies of "conflict." I. Variations in latency and strength of the rat's response in a discrimination-jumping situation. *J. comp. Psychol.*, 1941, **31**, 97–127.
3. Graham, C. H., & Gagné, R. N. The acquisition, extinction and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.
4. Haire, M. Some experimental data relevant to field and associative theories of discriminative learning. *J. gen. Psychol.*, 1942, **26**, 267–288.
5. Henmon, V. A. The time of perception as a measure of differences in sensations. *Arch. Phil. Psychol. sci. Meth.*, 1906, No. 8, Pp. 75.
6. Hilgard, E. R. An algebraic analysis of conditioned discrimination in man. *Psychol. Rev.*, 1938, **45**, 472–498.
7. Hovland, C. I. The generalization of conditioned responses. I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
8. Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
9. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
10. Humphreys, L. G. Generalization as a function of method of reinforcement. *J. exp. Psychol.*, 1939, **25**, 361–372.
11. Krechevsky, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1937, **44**, 97–133.
12. Lashley, K. S. An examination of the "continuity theory" as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26**, 241–266.
13. Pavlov, I. P. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* (Trans. by Anrep, G. V.) London; Oxford Univ. Press, 1927.
14. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
15. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century Co., 1938.
16. Solomon, R. L. Latency of response as a measure of learning in a "single-door" discrimination. *Amer. J. Psychol.*, 1943. (In press)
17. Spence, K. W. Continuous vs. non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.
18. Stevens, S. S., & Davis, H. *Hearing, its psychology and physiology.* New York: John Wiley & Sons, 1938.
19. Tolman, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, 1939, **46**, 318–336.
20. Tolman, E. C., & Minium, E. VTE in rats: overlearning and difficulty of discrimination. *J. comp. Psychol.*, 1942, **34**, 301–306.
21. Verplanck, W. S. The development of discrimination in a simple locomotor habit. *J. exp. Psychol.*, 1942, **31**, 441–464.
22. Weidensall, J. A critique of the discrimination test: a study in animal behavior. *Psychol. Bull.*, 1912, **9**, 57–59.
23. Williams, S. B., & Williams E. Barrier-frustration and extinction in instrumental learning. *Amer. J. Psychol.* (In press)

REMOTE ASSOCIATIONS AS A FUNCTION OF THE LENGTH OF INTERVAL BETWEEN LEARNING AND RECALL[1]

BY JOHN T. WILSON[2]

Washington, D. C.

Introduction

The following study is an investigation of remote associations as a function of the interval of time between practice at serial learning and a test for recall. The experiment was designed (*a*) to yield data on the occurrence and relative frequency of remote associations during recall after four different temporal intervals following learning, and (*b*) to determine whether or not remote associations dissipate with time in keeping with expectations based upon a theory of differential forgetting.

The relationship of the phenomenon under observation in this experiment to rote serial learning in general can be stated by defining remote associations as one type of intraserial phenomenon which occurs within the parts of a list during learning, but which is usually abstracted out of final reports dealing, for the most part, with the time, trials, and errors necessary to associate correctly the adjacent items in a list. To make the description more specific, let a serial list be represented by 1–2–3–4–5–6–7– . . . N, and let us assume that when it is being learned, the subject, under the usual instructions, is trying to associate 1 with 2, 2 with 3, 3 with 4, and so on to the end of the list. In other words, he is trying to form *immediately successive associations*, or *adjacent associations*, or *associations of zero degree remoteness* throughout the length of the list. These are not, however, the only associations the subject will form. Item 1 may lead sometimes to item 3, 4, 5 or 6; item 7 may lead to item 4, 5 or 3 as well as to 8. Such associations between non-adjacent items are termed *remote associations* and are further designated as either *forward* remote associations or *backward* remote associations, depending upon their direction from the item in the list which elicits them.

[1] This paper is a condensation of a thesis presented in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology, State University of Iowa, July, 1942.

[2] The writer wishes to express his gratitude to the late Prof. John A. McGeoch for his aid and advice in the direction of the study, and to Drs. Don Lewis and Benton J. Underwood for their criticism of the manuscript.

Evidence on the influence of varied temporal intervals upon the occurrence of remote associations is conflicting. Studies by Hall (2), Lepley (5) and Witmer (11) have shown no consistent results, and in no instance has the retention of remote associations been directly discussed.[3]

A priori there is no reason to believe that the 'forgetting' of remote associations cannot be included within the framework of a general theory of forgetting. However, one might infer that their differential rate of acquisition could have a counterpart in a differential rate of dissipation as compared to the forgetting of adjacent associations. Hull (4), Hovland (3), and Warden (10) have all formulated statements regarding differential rates of forgetting for inhibitory as compared with excitatory reaction tendencies, and McGeoch (8) has recently applied the concept of differential forgetting to certain phenomena from classical learning experimentation. The present problem is relevant to this latter application of the theory.

McGeoch (8, p. 142) points out that during the course of practice, a subject learns not only the correct responses but also incorrect and conflicting responses of which remote associations are examples. He hypothesizes that these conflicting response tendencies interfere with the appearance of the correct responses, but since they are reinforced to a lesser extent than are adjacent associations they are less well fixated, and being less well fixated, they suffer greater decrement with the passage of time than do adjacent associations.

Evidence which has been interpreted as favoring such a theory has been drawn from studies of reminiscence and distributed practice (8, pp. 142 ff. and 350 ff.). The appearance of reminiscence has been explained by postulating a more rapid dissipation of interfering responses during the interval preceding a measure of recall.[4] The greater efficacy of distributed over massed practice in learning is allegedly a function of the elimination of poorly fixated erroneous response tendencies during the interim between trials.

In the present experiment it was assumed that remote associations represent interfering response tendencies which are less well fixated than adjacent associations within a common list, and which are directly measurable by the association method (8, p. 74). Thus, the problem, along with investigating the occurrence and relative frequency of remote associations during recall at four different temporal intervals following learning, is to determine "whether remote associations are related to time interval in a way to fit the expectation

[3] For a complete discussion of remote associations, methods for studying them, theoretical and experimental papers relating to them, see McGeoch (8, Chap. 3).

[4] Such an explanation of reminiscence does not fit the evidence recently published by Melton and Stone (9). These experimenters report the absence of measurable reminiscence in learning material which had shown considerable anticipatory and perseverative errors during learning.

that they should be one of the forms of interference which is forgotten faster than are the right responses" (8, p. 147).

Procedure

Subjects.—The *S*s in the experiment were 48 college students, 27 women and 21 men, all of whom were enrolled in a course in elementary psychology. None had previously participated in an experiment in serial learning and all were naive regarding the field of experimental human learning. The nature and purpose of the experiment were not revealed to any *S* until the close of his or her last experimental session.

Learning Materials.—The learning materials consisted of 3 practice and 4 experimental lists of 16 two-syllable adjectives.[5] No list contained more than 2 adjectives with the *same* initial letter, and only 3 *pairs of similar* initial letters were permitted in any one list. No list contained more than 4 adjectives with the same terminal letter or syllable, and the frequencies of common adjectival endings were equalized as much as possible from list to list. Similar endings were not present in successive adjectives.

Randomized orders of the various lists, necessary in the recall test when the association method is used, were constructed by rearranging the serial order of each list. Only one random order was constructed for recall tests on each of the three practice days, while the lists for the four experimental days were randomized in three different serial orders which were equally divided among the 48 *S*s at the recall tests. The purpose of randomizing the serial order of the learned list for the recall test was to enhance the appearance of remote associations and to diminish the possibility for serial position associations built up during the learning of the list.

Procedure.—The basic procedure for measuring remote associations was that of the association method, adapted to the conditions of the present experiment. Four experimental days were preceded by three practice sessions, making a total of seven periods for each *S*. On no occasion was there more than one session a day for each *S*, and the sessions were arranged at or near the same hour each day. All learning was by the anticipation method with the adjectives being exposed serially on a Missouri-type memory drum. A 2-sec. exposure rate was used, with a 6-sec. rest pause between trials.

Three practice days served to familiarize the *S*s with all procedures that were to be employed in the experiment. Following the three practice days all *S*s served four experimental periods. On each of these four days the experimental list for that particular day was learned to a criterion of 12 or more correct anticipations on a single trial, and a test for associations was made after one of the four experimental time intervals.

The four intervals used were of 0.5-, 2.0-, 5.0- and 20.0-min. duration between the criterial trial and the test for recall. The 0.5-min. interval, which represents learning followed by immediate recall, was filled by the reading of the instructions for the recall test. The longer intervals were filled with color naming from a color card, up to 30 sec. before the test for recall, the 30 sec. being necessary for reading the recall instructions. After recall instructions, a recall test was made, using one of the random orders for the particular list learned that day.

In the recall tests on practice and experimental days, each adjective was exposed in the same manner as during learning except that the exposure during recall was continued until a response was made by the *S*. After a response was made, the next adjective in the random order list was immediately exposed by revolving the drum manually. It is to be noted that in the instructions for recall, emphasis was placed on the fact that the *S* should respond as quickly as possible with the first response that came to mind. Recall tests on the practice days served to make these instructions clear and there was rarely any delay in responding.

Controls.—Twelve *S*s, in addition to the 48 of the experiment, were tested with the association test lists, using a free-association technique, without any previous practice at learning the lists. This control was an attempt to verify the assumption that the responses being made in the recall tests were a function of the previous learning. It also served to determine the degree to which there was a tendency to repeat immediately preceding stimuli and responses during the recall tests. The various random orders of the recall test lists were again equally distributed over the group.

[5] These materials were previously used by Melton and Stone (9) in a study of reminiscence.

For complete counterbalancing with four experimental conditions, 24 permutations are required. Hence the 48 *Ss* were run as two groups of 24 each. Each experimental list was constant; that is, each *S* learned experimental list No. 1 on experimental day 1, experimental list No. 2 on experimental day 2, and so on, regardless of the particular condition for the day. This distributes equally over all conditions the practice effects which accrue with continued learning, as well as any unknown differences existing within the lists.

Scoring Method.—During the association tests the responses for each *S* were recorded. The adjacency or remoteness of each association made was determined by comparing it to the adjective which followed the stimulus adjective eliciting the association in the originally learned list. If the two were not identical, the degree and direction of the remoteness of the association made at recall was then determined by noting, in the list learned, the direction and number of items removed the response made was from the stimulus eliciting it.

Results

Control Data.—Inasmuch as the problem is to study remote associations as a function of the interval of time between learning and recall, it is essential that the results should not be a function of inequalities in the degree of learning prior to the introduction of the rest intervals. Table I reflects the efficiency of the counterbalancing

TABLE I

Measures of the Degree of Learning Prior to the Introduction of Rest Intervals

Condition	Mean Trials to Criterion of 12 Correct Anticipations	Mean Correct Anticipations on Criterial Trial
0.5 min. rest	9.48±0.62	12.90±0.32
2 min. rest	8.98±0.48	12.71±0.11
5 min. rest	9.29±0.42	12.96±0.15
20 min. rest	9.19±0.57	12.63±0.11

design which was used to equate original learning in the four conditions. Learning measures include (a) the mean number of trials required to attain the criterion of 12 correct anticipations on the criterial trial, and (*b*) the mean number of correct anticipations on the criterial trial. These data warrant the conclusion that there were no significant differences in the degree of learning of the lists preceding the interpolations of the various rest intervals.[6] The largest difference between the mean number of trials to learn to the criterion was 0.50 ± 0.66 [7] (C.R. = 0.76), between the conditions having the 0.5-min. and the 2.0-min. intervals. The largest difference between the mean number of correct anticipations on the cri-

[6] The differences between the means of all recall measures were tested for significance by using the ratios of the differences to the standard errors of the differences. The mean differences have been corrected for the correlation between the arrays (6, p. 135), and significance ratios of 2.02, 2.42, and 2.70 are minimal values for significance at the 5 percent, 2 percent, and 1 percent levels of confidence respectively (6, p. 240).

[7] All measures of reliability are standard errors of the means.

terial trial was 0.33 ± 0.20 (C.R. = 1.65), for the conditions having the 5.0-min. and the 20.0-min. intervals.

The data from the control group of 12 *Ss* verified the assumption that in the recall tests the responses being made were a function of prior learning and not merely repetitions of the immediately preceding stimulus or response items. From the 12 *Ss* tested for free associations, using the recall test lists, no free associations were given which were identical with stimulus words in the list, and only once was there a repetition of an immediately preceding response.

Types of Associations and Frequency of Each Type at Recall.—Three types of associations were possible at the recall test following an interpolated rest interval: (*a*) remote associations, (*b*) adjacent associations, and (*c*) free (non-list) associations. Of these three types, the remote and adjacent associations may be in either the forward or in the backward direction.

Table II shows the mean number of forward and backward remote associations which were made during recall following the four intervals. Remote associations occurred in both directions in all conditions. The forward remote associations outnumbered the backward approximately 4 to 3 in all conditions except following the 20.0-min. interval, where the advantage was in the same direction but less pronounced. As indicated in Table II, remote associations totalled about 50 percent of all associations made at recall in all four conditions.

TABLE II

MEAN NUMBER OF REMOTE ASSOCIATIONS AFTER THE FOUR INTERVALS OF TIME FOLLOWING LEARNING

Condition	Mean Forward Remote Associations at Recall	Mean Backward Remote Associations at Recall	Mean Remote Associations at Recall
0.5 min. rest	4.54±0.31	3.69±0.30	8.23±0.52
2 min. rest	4.54±0.34	3.65±0.29	8.19±0.49
5 min. rest	4.27±0.28	3.73±0.36	8.00±0.49
20 min. rest	3.83±0.30	3.29±0.31	7.13±0.50

The mean number of adjacent associations given following the various rest periods is shown in Table III. Again there is a predominance of associations in the forward as compared to the backward direction, the forward outnumbering the backward about 3 to 2 in all conditions. The total number of adjacent associations made was only about 40 percent of the total possible and the number of *forward* adjacent dropped from 75 percent at the criterial trial to about 22 percent at the recall test. This drop in "correct" associations is similar to that found by McGeoch (7) and Witmer (11) and

has been interpreted as resulting from the change in context which randomizing the recall order of the list produces.

The mean number of free associations at the time of recall is also shown in Table III. Such associations were infrequent following each interval, and they showed a slight tendency to increase as the interval increased in length.

TABLE III

MEAN NUMBER OF ADJACENT AND FREE ASSOCIATIONS AFTER THE FOUR INTERVALS OF TIME FOLLOWING LEARNING

Condition	Mean Forward Adjacent Associations at Recall	Mean Backward Adjacent Associations at Recall	Mean Adjacent Associations at Recall	Mean Free Associations at Recall
0.5 min. rest	3.63±0.41	2.52±0.29	6.15±0.52	1.63±0.33
2 min. rest	3.73±0.44	2.33±0.23	6.06±0.52	1.71±0.37
5 min. rest	3.60±0.35	2.38±0.29	5.98±0.50	1.96±0.40
20 min. rest	3.69±0.34	2.65±0.29	6.33±0.54	2.52±0.47

Effect of Increasing the Length of Interval.—Also to be seen from the data in Tables II and III is the effect of increasing the length of the interpolated interval upon the frequency of each type of response. The comparisons to be made in this instance are between the mean numbers of remote associations at recall following each of the four different rest periods, and between the mean numbers of adjacent associations made after these same intervals. In other words, such a comparison is in effect a description of the retention curves for remote and for adjacent associations under the present experimental conditions. To facilitate this comparison, the data of Tables II and III are presented in graphical form in Fig. 1.

Although primary interest is in the remote associations, a measure of the retention of adjacent associations was also made. The mean number of adjacent associations remains at a stable level over the four intervals.[8] The mean numbers recalled were 6.15, 6.06, 5.98 and 6.33 for the 0.5-, 2.0-, 5.0-, and 20.0-min. intervals respectively. The largest difference between these means was 0.35 ± 0.46 (C.R. = 0.76) for the 5.0- and 20.0-min. intervals.

[8] The absence of a drop in the retention curve of adjacent associations over increasing intervals of interpolated rest is of interest, especially in the face of the fact that Melton and Stone (9), using the same materials and similar conditions, do find a reduction in retention after the 20-min. interval. The present measure of recall differs however, in that the stimulus word order is randomized at recall rather than appearing in the original learning order. The items correctly recalled even after scrambling the list are thought to represent those highly overlearned items which still persist over the longest period in the Melton and Stone study. The items which serve to raise the recall score at the shorter intervals in the Melton and Stone study are presumably learned to a lesser degree and are eliminated at recall by the introduction of a random order of the originally learned list.

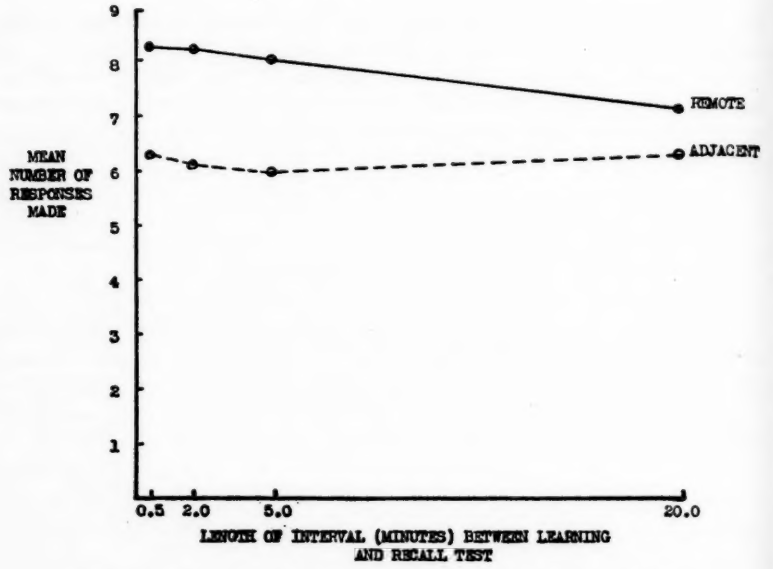

FIG. 1. The mean number of remote associations and adjacent associations after four intervals following learning.

The frequency of remote associations, while remaining at a fairly high level over all the intervals, did show a trend toward reduction with increasing lengths of interpolated rest. The average numbers of remote associations after the 0.5-, 2-, 5-, and 20-min. intervals were 8.23, 8.19, 8.00 and 7.13. The differences between the means of the shorter intervals are slight, the largest being 0.23 ± 0.47 between the 0.5- and 5-min. intervals (C.R. = 0.49). The differences between the mean numbers of remote associations at the 20-min. interval and at the three shorter intervals are larger, showing an increase from 0.87 ± 0.49 to 1.06 ± 0.52 and 1.10 ± 0.55 as the difference between the lengths of the intervals increases from 15 to 18 and 19.5 min. The significance ratios of these three differences between the means are 1.80, 2.04 and 2.00, respectively, and the levels of confidence, 7 percent, 5 percent and 5 percent.

DISCUSSION

The preceding data leave no doubt that under the present experimental conditions remote associations, both in the forward and backward direction, are present at recall following learning. These findings confirm similar evidence reported by Wohlgemuth (12),

Witmer (11) and McGeoch (7), all of whom employed the association method, and also support inferences for the existence of remote associations based upon the studies of Ebbinghaus (1), Hall (2), Lepley (5) and others using the method of derived lists. Although the ratio of forward remote to backward remote associations differed slightly from that reported by McGeoch and Witmer, this difference may be explained on the basis of the different criteria of learning employed. It is to be expected that as learning progresses toward the criterion of perfect forward reproduction (which was McGeoch's and Witmer's criterion) the ratio of forward to backward associations should increase if the list is to be finally learned.

With regard to the relationship between the length of interval following learning and the number of remote associations at recall, the preponderance of remote associations after the shortest interval following learning is not in keeping with the findings of Hall (2) and Lepley (5). This discrepancy is important with reference to the Lepley theory of remote associations. In his monograph, Lepley hypothesizes that remote associations are of the nature of trace conditioned responses. On the basis of this hypothesis he predicts that such associations should not function immediately or shortly after learning, but that after relatively longer periods of time they should be evident. McGeoch (8) has pointed out that Lepley's results, as well as the corroborating evidence from Hall concerning this prediction, may well be a function of their common experimental method and not a verification of the postulate. The present evidence for remote associations in both directions immediately following learning is supported by similar evidence reported by Witmer (11) and by McGeoch (7). These data are contradictory to Lepley's prediction and as such must be considered as evidence against his theory that remote associations are of the nature of trace conditioned responses.

With reference to a theory of differential forgetting as it has been applied to human learning by McGeoch (8), the data from the present study are not highly favorable. If such a theory is to be substantiated, remote associations, which supposedly represent one type of interfering response tendency within the list being learned, should decrease rapidly with an increased interval of interpolated rest. Furthermore, to support an explanation of reminiscence based upon this concept of differential forgetting, there should not only be a rapid decrease in the number of remote associations, but an accompanying increase in the number of adjacent associations recalled at some interval after learning.

Although there is a trend in the direction of satisfying the first of these two requirements, the decrease in the frequency of remote associations with each increase of time interval between learning and

recall is very small. The largest decrease can at best be interpreted as being only slightly significant. Should such a slight dissipation of remote associations with increasing intervals be interpreted as favoring a theory of differential rates of forgetting for remote as compared with adjacent associations, the evidence still does not support an explanation of reminiscence, for no reminiscence of adjacent associations was apparent after any interval. The remote associations which did dissipate were replaced by increasing numbers of free associations at recall following the longer intervals. A similar absence of reminiscence has been reported by Melton and Stone (9) in a study employing the same materials and similar experimental conditions. Whether such findings reflect a shortcoming of the theory or are merely the product of conditions and materials common to both studies only further experimentation will tell.

Summary

Forty-eight subjects learned serial lists of 16 two-syllable adjectives at a 2-sec. exposure rate to a criterion of 12 correct anticipations. Following the criterial trial, rest intervals of 0.5, 2, 5, and 20 min. were interpolated after which recall of the learned lists was tested by the association method.

The purpose of this procedure was (*a*) to measure the occurrence and frequency of remote associations after four different intervals following learning, and (*b*) to determine whether or not these remote associations dissipate with increasing lengths of interpolated interval in keeping with expectations based upon a theory of differential forgetting for remote associations as compared with adjacent associations.

On the basis of the data derived from this experimental procedure it was found that:

1. Remote associations of all degrees of remoteness in both the backward and forward directions were present at recall after all four intervals following learning.

2. Forward remote associations outnumbered the backward remote associations approximately 4 to 3 in all conditions except that following the 20-min. interval, where the advantage was in the same direction but less pronounced.

3. Remote associations did not dissipate with increased lengths of interpolated intervals in keeping with expectations based upon a theory of differential rates of forgetting for remote associations as compared with adjacent associations.

(Manuscript received April 30, 1943)

References

1. Ebbinghaus, H. *Memory: a contribution to experimental psychology.* (Trans. by H. A. Ruger and C. E. Bussenius.) New York: Teachers College, Columbia University, 1913. Pp. viii + 123.
2. Hall, M. E. Remote associative tendencies in serial learning. *J. exp. Psychol.*, 1928, **11**, 65–76.
3. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
4. Hull, C. L. The conflicting psychologies of learning—a way out. *Psychol. Rev.*, 1935, **42**, 491–516.
5. Lepley, W. M. Serial reactions considered as conditioned reactions. *Psychol. Monogr.*, 1934, **46**, No. 205. Pp. 56.
6. Lindquist, E. F. *A first course in statistics.* (Rev. ed.) New York: Houghton Mifflin, 1942. Pp. vii + 242.
7. McGeoch, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.
8. McGeoch, J. A. *The psychology of human learning: an introduction.* New York: Longmans, Green, 1942. Pp. vii + 613.
9. Melton, A. W., & Stone, G. R. Retention of serial lists of adjectives. *J. exp. Psychol.*, 1942, **30**, 295–310.
10. Warden, C. J. The distribution of practice in animal learning. *Comp. Psychol. Monogr.*, 1923, **1**, No. 3. Pp. 64.
11. Witmer, L. R. Remote associations in serial learning. Ph.D. Dissertation, Yale University Library, 1935.
12. Wohlgemuth, A. On memory and the direction of associations. *Brit. J. Psychol.*, 1913, **5**, 447–465.

GENERALIZATION IN THE INITIAL STAGES OF LEARNING NONSENSE SYLLABLES: I. INTEGRAL RESPONSES

BY B. R. PHILIP

Fordham University

AND

HELEN E. PEIXOTTO

Hunter College

Recently attempts have been made to relate various facts of verbal learning to characteristics of the conditioned response, particularly to the concept of generalization. Experimental studies of generalization have been quite numerous in the field of the conditioned response, but generalization of verbal material has not been so thoroughly investigated; and those investigations which have been made are on generalization of stimulus, as described by Hull (**5**), Gibson (**1**, **2**, **3**) and several others. It has been noted that generalization of response also takes place in the learning situation (**4**). This paper is a study of generalization of stimulus and of generalization of response in the initial stages of learning difficult nonsense syllables. It also studies the discrimination of stimuli and the differentiation of responses in the same situation.

Since this study utilizes lists of nonsense syllables where every learning trial is followed by a written reproduction, an opportunity is also offered of tracing the development of learning as indicated by the progressive changes in the responses. Studies of the learning of verbal material have chiefly concerned themselves with correctly reproduced responses, and imperfect responses have as a rule been neglected. In the present study the development of imperfect responses will be studied as well as that of perfect responses because it is felt that the former also show a form of learning.

Definition of Terms

Since terms ordinarily used in conditioned response studies are here applied to verbal material, it is of importance that these terms be defined, as they apply to this experiment.

In the normal paired associate procedure the response member must not only be properly reproduced (*i.e.*, spelled), but it must be

correctly located in reference to the proper stimulus syllable. In the conditioned response terminology, when this occurs the response is differentiated and there is discrimination of the stimulus. By generalization is meant the assigning of a response to an inappropriate stimulus; *i.e.*, the syllable, though correctly reproduced, is incorrectly localized.

Hitherto we have been referring to syllables correctly reproduced, whether or not they are correctly localized. But responses may not be correctly spelled, and yet indicate some degree of learning. Hence, a distinction is made in this paper between integral and partial responses, according as the syllables are correctly or incorrectly reproduced. Depending upon the correctness of their localization, partial responses may show a form of discrimination and of generalization, which may be of importance in the initial stages of learning.

Thus far the emphasis has been on the response; now it is proposed to consider the stimulus. The assigning of a correct response to a particular stimulus is called *discrimination;* obviously this occurs for both stimulus and response since both members of the paired associates are correctly localized. When the pairing of stimulus and response is not correct there is *generalization*, and generalization of response naturally entails a corresponding generalization of stimulus (1). But it is important to note that these are not identical, for they differ in the locus of the generalization, as is indicated in the following diagram. In *generalization of stimulus*, the reference is to all the incorrect stimuli (S_1, S_2, S_3, . . . S_k) that evoke one particular response (R_a); in *generalization of response* the reference is to all the incorrect responses (R_1, R_2, R_3 . . . R_n) that are evoked by one particular stimulus (S_b).

Since the stimuli under the conditions of this experiment are necessarily fixed in location, both generalization of response and generalization of stimulus are determined by the responses. As the stimuli are presented to the subject, although in a constantly varying order from recitation trial to recitation trial, the subject may or may not allocate an appropriate response to them, but he cannot vary the stimuli either in reproduction or localization, and therefore he cannot generalize them in the more restricted meaning of the term. In such

cases generalization of stimulus is to be inferred from the mode of the response.

Accordingly, the following terms are defined as they apply to this study. Again it must be emphasized that the correctness of the written reproduction of a response syllable when the paired associate technique is used depends both upon its reproduction and its location.

In this study we will refer to correctly reproduced responses, whether properly or improperly localized in reference to their stimulus, as *integral responses*. The proper localization of integral responses is called *discrimination*, and this term is applied to both stimulus and response, although in the latter case it is often called *differentiation*. The improper localization of integral responses is called *generalization of response*, when the emphasis is laid upon all the incorrect integral responses evoked by a particular stimulus. The pairing of an integral response to an inappropriate stimulus is called *generalization of stimulus*, when the emphasis is upon all the incorrect stimuli that evoke one particular response.

Incorrectly reproduced responses are called *partial responses* when one letter only of the syllable is wrong, or the syllable is inverted. The response syllable may be so modified or may be so incorrectly spelled as to be impossible of identification with any of the response syllables of the paired associates; this type is called *inadequate response*. Partial responses may be either correctly localized or incorrectly localized; in the first case the result is *partial discrimination* (*both* of the stimulus and response), and in the second case, *partial generalization* (*either* of the stimulus *or* the response). Inadequate responses cannot be localized with reference to any particular stimulus.

This paper deals only with what has been defined as integral responses. The results obtained from the partial and inadequate responses will be presented in a subsequent paper.

Purpose

The purpose of this paper is to study generalization and discrimination in the learning of nonsense syllables, particularly in relation to the following topics:

1. The relative frequency of the various types of response. It seems of importance to ascertain the relative proportions of generalized and of discriminated responses, and also the proportions of integral, partial and inadequate responses, in the initial stages of learning. In this case the emphasis is upon the total responses yielded by the subjects during the learning situation.

2. The variation of the relative frequency of occurrence of the various types of responses as learning proceeds. Here the emphasis

is upon the responses from trial to trial, or at different stages of the learning process. This topic has received the attention of a number of investigators, who have been interested in the form of the curve of generalization with learning. The material for this section will be found in the second paper which deals with partial and inadequate responses.

3. Gradients of generalization. Do gradients of generalization along a dimension of serial distance from the discriminated response occur in the learning of nonsense syllables? In other terms does frequency of generalization diminish in proportion to serial distance from the basic syllable, *i.e.*, the correct response to a given stimulus? Such gradients are commonly found along appropriate dimensions in conditioned response studies. In studies of verbal material they have been noted along a dimension of similarity (**2**). The phenomenon of spread of effect consistently found by Thorndike and others is of essentially the same nature as what is here called the gradient of generalization (**9**).

4. The relationship between discrimination and generalization for the individual pairs of associates, independent of serial position. In determining gradients of generalization the emphasis is upon serial position; here the emphasis is upon each particular syllable, no matter what its serial position. As we have been differentiating between generalization of response and generalization of stimulus it is of interest to determine whether the individual syllable yields different values for these two variables, and to see how both change with discrimination values for the same syllable.

Procedure

The usual classroom situation was used for a group experiment, 504 high school students divided into approximately equal groups acting as subjects. Five lists of syllables, consisting of 4, 6, 8, 10 and 12 paired associates, were presented by film projection. In all, 16 different pairs of syllables were used, the same pairs being used in all the series except the 4-series, to allow for interserial rotation of the pairs in the learning order. Each group was given the 4-series and a major series (*i.e.*, one of the series containing 6, 8, 10 or 12 pairs). Thus, for every group, the 4-series preceded the major series. This procedure served two purposes: it gave a practice exercise to the subjects, and yielded a measure of the comparability of the various groups. The paired associates appeared on the screen one pair at a time in the learning trials. In the recitation trials the first (stimulus) syllable only of each pair appeared, one syllable being on the screen at a time, and the subjects were required to write the second (response) syllable which corresponded to the given stimulus syllable. After each learning trial a recitation trial was presented. The number of learning and recitation trials per series was restricted to one and a half times the number of pairs in the series, since complete learning by all subjects was not desired. More complete details of procedure may be found in a previous article (**8**). As the emphasis was on intraserial effects in relation to learning, the following experimental precautions were observed: 1. The syllables to be learned were of low associative value. 2. The learning order of the paired associates in each series was kept constant throughout the experiment. 3. The recitation order was randomly varied for each trial.

The method of determining generalization and discrimination was quite simple. It consisted in analyzing the lists of reproduced responses for correctness of reproduction and of localization

within the series. Accordingly, tabulations were made and analyzed for every recitation trial of all the responses to each stimulus, and of all the stimuli which evoked each particular response, due regard being given to the serial location of every stimulus and response.

It must be emphasized that this paper deals chiefly with the initial stages of learning, to which the findings and interpretations are restricted. With difficult material of this type, in which the learning by unpracticed subjects was far from complete, the findings, as might be expected, differ from those of experimenters who have used easier material and who have carried the learning to a more advanced degree.

Results

Relative Frequency of Occurrence of Various Types of Responses

The relative frequency of the various types of response is given in Table I, for each of the four series. The values shown are in

TABLE I

Percentages of Occurrence of Various Types of Response for Each Series

		6-Series	8-Series	10-Series	12-Series
Integral	Discrimination	8.5	8.7	7.6	7.2
	Generalization	14.1	7.9	6.3	7.0
Partial	Discrimination	4.8	4.8	3.7	2.8
	Generalization	9.4	9.0	8.4	7.8
Inadequate		63.2	69.6	74.0	75.2
		100.0	100.0	100.0	100.0

terms of the percentage of the total possible responses for each series. The percentages are fairly consistent from series to series; the inadequate responses increase in percentage from the shorter to the longer series whereas the other values diminish in proportion.[1] As Table I is based upon the responses totalled for all trials, which differ in number from series to series, there is no possibility of comparing from this table the degree of learning evidenced in the various series. The variation of the responses with degree of learning will be discussed in a subsequent paper. Since learning in this experiment has not progressed beyond the initial stages, percentages of inadequate responses, as may be expected, are considerably higher than those of the others. This fact may indicate that the majority of the syllables are not learned to any degree; it may also imply that the subjects concentrate upon perfecting the learning of the syllables they have partially learned. According as they learn these syllables they select additional ones for learning. It is of interest to note that in this experiment generalized responses, both integral and partial, occur

[1] The two types of inadequate response seem to be given about equally often. Thus in the 6-series, 45 percent of the inadequate responses were incorrectly reproduced, and 50 percent were omitted. Which type of response is given seems to depend upon the individual preference of the subject.

more frequently than do the discriminated responses, and that the integral responses are more frequent than the partial responses. This fact also may be explained by assuming that the subject selects certain syllables, and in general learns them sufficiently well to reproduce them correctly as integral responses, but their correct localization demands further learning which comes only in later trials.

Generalization According to Serial Position

In order to determine a generalization gradient it is necessary to have a base or reference point. In this experiment the gradient is determined in serial distance from the position of the basic syllable. The basic syllable, (the discriminated response), is the correct response syllable corresponding to a particular stimulus in the fixed learning order of each series. Obviously the base changes for each pair of syllables within a series. When the response syllable is correctly given the gradient distance from the base is zero. All integral responses incorrectly located indicate generalization. If to a certain stimulus a response syllable three places away from the corresponding basic syllable is given, the gradient distance is three. Since the recitation order is constantly changing, the gradients are determined in terms of the syllable positions in the fixed learning order, and naturally a separate gradient must be determined for each pair of syllables in a series. Series 6 is given in Table II by way of ex-

TABLE II

INTEGRAL RESPONSES FOR THE 6-SERIES

TOTAL SCORES FOR STIMULUS AND RESPONSE SYLLABLES

The stimuli are given in the first column and the responses in the first row, in the original learning order. When a syllable is given as a stimulus or a response it is capitalized, and its paired associate appears in lower case letters. This table is read as follows: To the stimulus ZOF, the correct response GEX was given 45 times; the incorrect responses WUQ and VAF were given 33 and 17 times respectively, etc. In like manner, reading down the column, the response GEX was given 45 times to the appropriate stimulus ZOF, and 32 and 43 times to the incorrect stimuli YAV and GEC respectively, etc.

6-S	zof GEX	yav WUQ	gec VAF	zuv CIJ	qos MEF	yix ZOJ	Tot.
ZOF gex	45	33	17	35	46	26	202
YAV wuq	32	33	29	45	57	23	219
GEC vaf	43	38	69	31	59	51	251
ZUV cij	30	29	24	43	43	32	201
QOS mef	19	12	18	39	339	18	445
YIX zoj	44	21	27	56	50	69	267
Tot.	213	116	184	249	594	179	1585

emplification, and only integral responses are there tabulated. The first step in determining the gradient consists in the listing for every trial of the number of times each integral response syllable was assigned to each stimulus. These listings are then totalled for all the trials as in Table II.

In the second step the procedure as exemplified for Series 6, in Table III, is as follows: Starting with the last row of Table II, the number of occurrences of the correct syllable (69 for the pair YIX—ZOJ in this instance) is placed under the column headed zero, or basic column, and the number of incorrect integral responses is placed in its appropriate column before or after the basic syllable according to the serial distance of those responses from their base. In like manner the second last row for the pair QOS—MEF and all the rows up to the first row are located. Thus all the discriminated responses will appear in the basic column and there will be tabulations of generalized responses on both sides of the basic position. Since the gradients are approximately the same on either side of the basic syllable, the totals from corresponding distances on either side of the base are then added together and averaged in accordance with the number of listings in each column, to form the generalization gradient which appears in the last line of Table III. It is to

TABLE III

GENERALIZATION GRADIENT

6-Series

This table shows, for the 6-series, how the integral generalization gradient is derived from Table II. Distance is estimated from the basic syllable, the correct syllable response to a given stimulus based on the learning order. The first row of this table is taken from the row opposite the pair YIX—zoj in Table II. The other rows are taken in ascending order from Table II, and are placed in their appropriate columns, in reference to the basic syllable. The totals are then averaged for frequency of occurrence.

Serial Distances	5	4	3	2	1	0	1	2	3	4	5
	44	21	27	56	50	69					
		19	12	18	39	339	18				
			30	29	24	43	43	32			
				43	38	69	31	59	11		
					32	33	29	45	57	23	
						45	33	17	35	46	26
Sub-totals	44	40	69	146	183	598	154	153	103	69	26
Totals						598	337	299	172	109	70
Averages						98.0	33.7	37.4	28.7	27.5	35.0

be noted that the generalization gradient is determined from more sub-totals closest to the base, whereas the gradient value for the most remote position in each series is determined from only two sub-totals. Hence the reliability of the gradient diminishes with its distance from

the base. When the generalization gradient is thus determined from the averages for all the serial positions in the series, the gradient is the same for both stimulus and response. That is, in this case, as in every case where the generalization values for the individual syllables are not separated, the values for both stimulus and response are identical. As was pointed out above, generalization of response entails generalization of stimulus, so that when merely position is concerned the generalization values for stimulus and response will be the same—they can only vary when the individual syllables are studied.

Fig. 1 presents the gradient of generalization for serial position. The correct integral responses or the discriminated responses are

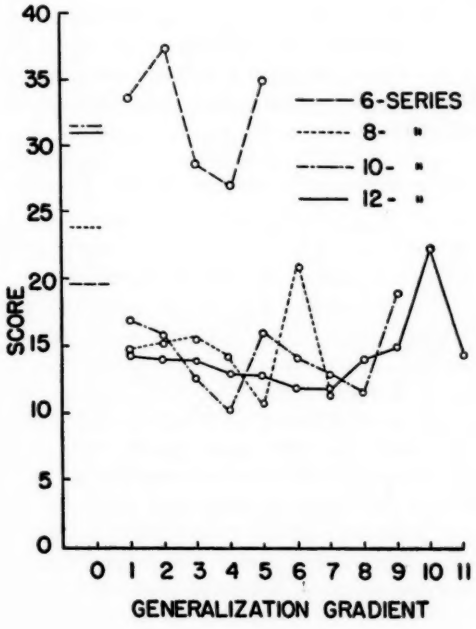

FIG. 1. Generalization gradient according to serial position of the syllables in the learning order, for integral values.

The ordinates indicate the total score. The abscissae give the generalization gradient according to distance from the basic (discriminated) syllable.

Above position 0, the lines, coded according to series, indicate discriminated values, the ordinate scale for which is one-fifth that of the generalized values.

Gradients may be seen at the beginning of each series; the values for the terminal positions of each series are probably not reliable.

indicated by the coded line above serial position zero; their ordinate scale is one-fifth of that of the other values. The generalization gradients start at position one. It is apparent that series 12, for the first seven positions, and series 10 for the first four positions, show a slight but fairly smooth gradient. Series 8 up to the fifth position, and series 6, up to the fourth position, likewise show a gradient, although with slight initial inversions. It is to be expected that the positions towards the end of each series would fluctuate rather widely in value owing to the lesser reliability of these values, since they were averaged from fewer listings and are more subject to the effect of associational and finality factors (**8**). As might be expected, there is some indication that the shorter series have steeper gradients because the localization of the integral syllables is naturally more precise in the shorter lists. The initial inversions for the 6-series and the 8-series are probably artifacts due to the wide differences in actual association value of the syllables. In both these series the first position in the generalization gradient happens to have syllables of much lower associative values than has the succeeding positions in the immediate vicinity.

Relationship Between Generalization and Discrimination

In the gradient of generalization the emphasis is upon positional effects, and the data are differentiated according to positional distance from the basic syllable. In the material to follow, generalization is independent of position but is differentiated according to the individual syllables. Hence, in the following section a different mode of calculating generalization was followed. The procedure consisted primarily in subtracting the total number of discriminated responses from the total number of integral responses for every syllable. All values were obtained from tables similar to Table II, in which the total integral responses corresponding to each stimulus syllable are given in the Total column, and the total number of times each response syllable was given as an integral response to all the stimuli is listed in the Total row. The correctly localized or discriminated responses are selected from their appropriate position in each row.

The manner of obtaining the values for generalization of response is here exemplified using the pair QOS—MEF. In Table II, opposite this pair of syllables is listed the number of times the stimulus QOS elicited the various integral response syllables; under the Total column it is seen that this syllable was responded to, in all, 445 times, but only 339 times by the correct or discriminated syllable MEF. Consequently, 106 (*i.e.*, 445 − 339) is the total number of integral generalizations of response in series 6, to the stimulus QOS. The

generalization of response for this syllable must be determined from all four series in which it appeared as stimulus, and averaged for the number of trials. The manner in which this calculation was carried out may be seen from Table IV. To obtain generalization of stimulus the same procedure was followed except that the totals in the various *rows* are substituted for those of the *columns*.

It was necessary to average for number of trials because only six of the syllable-pairs appeared in all four series, the other pairs appearing in three, two or one series. The corresponding numbers of trials in which these syllables appeared were therefore 54, 45, 33 and 18. It is to be noted that the discrimination and generalization values listed in Table V according to rank order of discrimination are not averaged for the number of *S*s in each of the groups. Since the groups contained approximately the same number of *S*s, the relative values for discrimination and generalization are but slightly changed when corrected for varying number of *S*s, and the trend of the results is in no way affected.

The average values per trial, similar to those which appear in the last line of Table IV, were computed for all the syllables, and appear in Table V. Although no special attempt was made to keep the

TABLE IV

DETERMINATION OF GENERALIZATION OF RESPONSE FOR QOS—MEF

This table shows how the discrimination and generalization scores of Table V were obtained, based upon averages per trial.

Series	Number of Trials	Integral Responses	Discriminated Responses	Generalized Responses
6-S	9	445	339	106
8-S	12	412	349	63
10-S	15	393	283	110
12-S	18	513	365	148
Totals	54	1763	1336	427
Average per trial		32.1	24.7	7.9

degree of learning constant for the different series, other than having the number of trials one and a half times the number of syllables in a series, degree of learning did turn out to be substantially the same in all four series, as will be shown in a subsequent paper. Worthy of note in Table V is the wide range in discrimination values. Though the syllables were selected from lists of lowest association values according to Glaze, they are far from equal in discrimination values, as was pointed out in the previous section and also in a previous paper (**8**).

To show the relationship between generalization and discrimination Fig. 2 is presented. The syllables are arranged along the

TABLE V

RELATION BETWEEN DISCRIMINATION AND INTEGRAL GENERALIZATION

The values are given for syllables listed in descending order of discrimination. The frequency of occurrence of the syllable pair is given in column 2; thus QOS—MEF appeared in all four series. The values are averaged for frequency of occurrence.

Syllables	Occurrence	Discrimination	Generalization of Response	Generalization of Stimulus
QOS—MEF	4	24.7	7.9	14.3
QIM—JEX	2	23.8	7.3	21.4
GEC—VAF	4	16.9	10.2	9.6
MIB—QUJ	3	13.8	11.2	9.9
QOJ—CEF	1	10.2	10.5	10.3
ZOF—GEX	4	7.9	9.4	14.2
KOJ—YUF	1	3.9	9.6	6.3
ZUV—CIJ	4	3.9	9.7	12.2
YAV—WUQ	4	3.4	10.2	7.3
YIX—ZOJ	4	2.6	11.0	4.0
VUF—ZEQ	3	1.7	9.7	2.1
YIF—VUQ	2	1.6	8.5	3.2

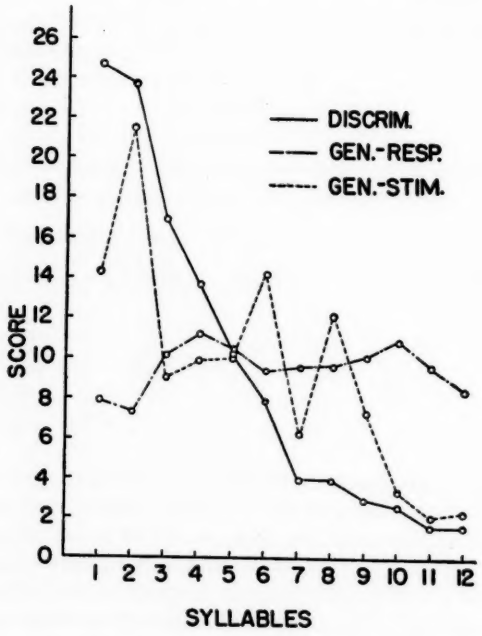

FIG. 2. Relation between generalization and discrimination for integral responses.

The ordinates are the total scores, averaged for number of occurrences of the syllables in the four series.

The abscissae give the order of discrimination; the most frequently discriminated syllables are first in order.

abscissae according to rank order of discrimination. The corresponding generalization curves, both of stimulus and of response, are placed on the same figure. It may be seen that the generalization of response for each syllable is largely independent of its discrimination, but that, broadly speaking, generalization of stimulus varies directly with discrimination. Some clue to possible explanations of this finding may be gleaned from the fact that experimenters have found that the stimulus member of a pair is more susceptible to inhibition than is the response member. Thus McGeoch states, "Community of stimulus members is much more inhibitory than community of second members" (**6**, p. 493). Since generalization has been found by Gibson (**3**) to be associated with differentiation or distinguishability, and since in all probability those syllables which are best distinguished are best discriminated in the sense used in this paper, one would expect to find generalization to vary with discrimination. Gibson's findings, however, apply only to generalization of stimulus. The response members of the pair, being less subject to inhibitory influences than the stimulus members, would not be expected to show the same degree of variability in generalization, and hence their generalization values would not vary with the discrimination values to the same degree as those of the stimulus members.

Discussion

The concepts of generalization and of discrimination have not received as much attention in discussions of the higher mental processes as they have in the field of conditioning. That these processes as they occur in verbal learning are in need of clarification is indicated by the importance that they have been found to contribute to conditioned learning, in which field they have been carefully distinguished. Thus, Hilgard and Marquis (**4**, p. 176) make the following distinctions: "The partial equivalence of different stimuli in evoking a conditioned response is known as *sensory generalization*. There is also a degree of equivalence among the responses which may be called *response generalization*, so that a stimulus which has come through training to elicit a particular response, may under some circumstances, come to elicit a different response without special training." These authors also distinguish between discrimination of the stimuli and discrimination of the response, both of which they refer to as selective reactions.

There is some indication that experimenters on generalization in the ideational field realize that distinction should be made between the various types of generalization. They do not consistently differentiate, however, between generalization of concepts, generaliza-

tion of sensory stimuli which the subject must learn to discriminate, and generalization of responses which the subject must differentiate. In defining generalization, Gibson (1, p. 201) is apparently referring to generalization of stimulus. She recognizes the distinction, however, since she says: "It is conceivable that something analogous to generalization of stimulus items might occur among response items, especially since every response is in a sense a stimulus as well." It must be noted that in this paper we have been dealing with a different type of generalization than that which is the principal object of Gibson's study. She, however, identifies the type here studied, and calls it 'multiple generalization.' Her results show a 'multiple generalization' of ten percent, which is in accord with the findings of the present experiment, as may be seen from row two, Table I, which lists the percentage of integral generalizations of response.

It is of interest to compare the classification of responses given in this paper with Melton and Irwin's (7) classification of overt intrusions in their study of specific interlist interference effects, although the wide differences between the experimental procedures of the two studies obviate any direct comparison. Their classification was fourfold: intrusions based upon (*a*) serial position, (*b*) formal similarity, (*c*) serial position and formal similarity, and (*d*) unclassified intrusions. When tabulating intrusions based upon serial position, they took into account only those not more than one serial position removed from the corresponding correct syllable. In some respects this type would correspond to our generalized response, though we took all serial positions into account. They point out the difficulty of precise determination of the similarity factor responsible for specific intrusions, yet they do not differentiate between identical intrusions and similar intrusions (the integral and partial responses of this study). They find that the frequency of the intrusions is a function of the degree of interpolated learning, that the maximal number of intrusions based upon serial position occurs near the beginning of the learning and that maximal retroactive inhibition occurs at the same stage (7). These findings are related to Gibson's finding that a maximal point of generalization occurs fairly early in learning (2). As will be shown in a subsequent paper, we also find that generalization of response is a function of the degree of learning, and reaches its maximum before the end of the stage of initial learning to which this experiment was carried.

The role played by formal similarity in generalization could not be adequately determined in this study. For one thing, when the paired associate method is used, it is impossible to determine whether the response given is based upon similarity with the stimulus syllable or with the correct response syllable. Generalization based

upon serial position is susceptible of precise determination, and, as this study indicates, is a factor that is important at least in the early stages of learning.

(Manuscript received April 21, 1943)

Bibliography

1. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
2. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
3. Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–201.
4. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century Co., 1940, xi + 429 pp.
5. Hull, C. L. Quantitative aspects of the evolution of concepts. An experimental study. *Psychol. Monogr.*, 1920, **28**, No. 123.
6. McGeoch, J. A. *The psychology of human learning.* N. Y.: Longmans, Green & Co., 1942.
7. Melton, A. W., & Irwin, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 170–204.
8. Peixotto, H. E. Intraserial inhibition as measured by reproduction. *J. exp. Psychol.*, 1942, **31**, 17–35.
9. Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contr. Educ.*, 1933, No. 580, 72 pp.

STUDIES IN OBJECT-PREFERENCES. I. THE EFFECT OF TEMPORAL PROXIMITY

BY FRANCIS W. IRWIN, FANNIE M. ARMITT, AND CHARLES W. SIMON

University of Pennsylvania

In an interesting series of experiments, Herbert F. Wright (2) has studied the choices made by children and adults among a variety of objects which presented different amounts of difficulty of attainment. This work was prompted in part by the existence among many peoples of proverbs of which "grass on the other side of the fence is greener" may be considered to be representative. Sayings to the contrary effect appear to be almost non-existent. Since, in those experiments which seemed to him to be free from disturbing influences, Wright's subjects tended to choose objects relatively difficult to attain, whether the preference was expressed verbally or otherwise, he concluded, in Lewinian language, that a barrier enhances a positive valence. The high degree of generality claimed for this conclusion was supported by favorable results in situations ranging from children's choices among sticks of candy in a laboratory situation to adults' preferences for pies in a cafeteria.

Wright's results conflict with certain findings of Fajans (1) for young children. Using as a criterion the relative amount of time spent in activity directed toward an attractive object such as a doll or rattle, Fajans concluded that in the case of infants six months to one year old the attractiveness of objects decreased as the distance of the object from the infant increased. This was true over a range of distances from 9 to 100 cm. With older children, up to about six years of age, there appeared to be no difference in attractiveness with increase of distance. Fajans attributed the lack of consistency at the different ages at least in part to the fact that the older children perceived the more distant objects to be impossible to obtain by direct action, while the younger ones did not.

In most of his experiments Wright either varied the spatial distance of objects from the subject, or interposed some physical object between the subject and the goal-object. His definition of 'barrier' was general enough, however, to permit him to use temporal distances as well; and those few experiments in which they were used produced results which he believed supported his general conclusion that a positive valence is enhanced by the presence of a barrier. The experiments about to be described were intended

originally to constitute a further investigation of the effect of a temporal barrier upon the attractiveness of objects. As will be seen, however, their outcome suggests that variations of temporal distances are not entirely analogous, psychologically, to variations of physical distances.

All of the experiments were performed with children ranging in age from 4 to 9 years. The procedures had in common determining the preference of each child between two toys, one of which the child was to receive as his own possession almost immediately, and the other of which he would possess after a longer interval. The time between the obtaining of the one toy and the obtaining of the other was intended to serve as the temporal barrier, the effects of which upon the children's preferences were to be studied. It is important to note that an effort was made to have each child understand that he would eventually receive *both* toys, so that his expression of preference did not constitute a decision as to which of the two toys he was to possess.

Experiment IA

Subjects and Procedure

The subjects were 25 children with a mean age at their last birthday of 5.7 years; at ages 4, 5, 6 and 7 years there were, respectively, 2, 8, 10 and 5 cases. The group contained 16 boys and 9 girls.[1] All cases were observed singly.

The toys used were bought at Woolworth's and other stores for 5 or 10 cents apiece. Many of the children in this and the other experiments were 'underprivileged'; and, in general, the subjects seldom assumed an attitude of indifference to the toys. The toys were varied considerably from one child to another, with attention to the age and sex of the child. The experimenter attempted to present before a child two toys of approximately equal attractiveness; but, in addition, formal control of this factor was achieved by using each pair of toys at least twice (or some even number of times), one toy being the more accessible toy in one case and the less accessible in the other. (The odd number of subjects caused one exception to this rule.) Among the toys used were: bubble set, hair ribbon, soldier, anti-aircraft truck, marbles, auto, knife, notebook, etc. The experimenter took care that toys which were promised to the children for a later date were in fact delivered to them.

The procedure itself was simple. After establishing rapport with the child, the experimenter put two toys before him and said:

"Here are two toys. You will be given [the one named] today. A week from today you will be given [the other named]. Which do you like the better?" When the child had expressed a preference, he was given the toy which he had been promised for 'today.' After he had answered the question, "Does a week seem like a long time to you?" he was dismissed.

The experimenter in this procedure was F. M. A.

Results and Discussion

Of the 25 children, 22 expressed a preference for the toy which they were about to receive, while only 3 preferred the toy which was promised for one week later. On the hypothesis of an equal

[1] We wish to express our appreciation to the University of Pennsylvania Settlement House, the Bedford Center, and the Seaford (Delaware) Public School, through whose coöperation we were able to obtain the subjects of this and the later experiments.

division of preferences by chance, the expected frequency of preferences for either of the two alternatives would be 12.5 with a standard error of 2.5. The obtained frequency varies from expection by 3.8 times the standard error, and it is therefore highly improbable that such results would be obtained by chance.

The interval of a week was chosen, not under the supposition that all of the children would have similar and definite notions of the length of such a period of time, but in the expectation that it would at any rate appear to the child to be a relatively *long* time. The children's answers to the question concerning the length of a week supported this expectation, but gave the experimenter the impression that the subjects had in general no clear conception of a week as a period of time.

It was suggested to us that in this experiment a child who indicated verbally a preference for the toy which was to be received the sooner might in fact be saying, as it were, "I should rather have a toy now than have to wait for one." That is, the child might be supposed to have misunderstood the instructions in such a way as to lead him to believe that his statement of preference would determine which toy would be given to him. Although nothing was observed which tended to support this objection, there was also no direct evidence to refute it. Therefore, in order to avoid this possibility, and also to attempt to confirm these first results, a procedure was developed in which the preferences were expressed by behavior other than verbal, and in which there appeared to be a smaller likelihood of misunderstanding on the part of the child.

Experiment IB

Subjects and Procedure

The subjects in this experiment were 25 children in the Public School of Seaford, Del., of whom 13 were boys and 12 were girls. Their mean age was 6.0, there being 1 at age 7, 22 at age 6, and 2 at age 5.

The procedure was in general similar to that of Experiment IA. The experimenter set two toys about a foot apart on a table before which the child stood, and said:

"Do you see these two toys? Well, I'm going to give you this one [pointing] in just a few minutes, as soon as I've finished what I'm doing. [The experimenter appeared to have writing to do.] And I am going to give you this one [pointing] a week from now. Do you understand?" No matter what the child replied, these instructions were repeated. *E* then pointed to one of the toys and said, "Now tell me, when are you going to get this one? And when am I going to give you this one [pointing to the other]?" If the child failed to reply correctly, the first instructions were repeated, and so on until the child could state correctly when he would receive each toy. *E* then said, "While I am working here, you may take either toy that you want to over to that chair there and play with it." Before the child expressed his decision, *E* added, "Do you understand? You are going to receive this toy [pointing] in a few minutes *anyway*, and you are going to be given this toy [pointing] a week from now *anyway*, but right now, right at this moment while I am working here, you may take *either* toy over in that chair and play with it [or, 'look at it']. Which do you want to take?"

This temporary suspension of the barrier was designed to permit a behavioral choice. It also was intended to make it less likely that the preference so expressed could be interpreted as meaning merely that the child did not wish to have to wait for a toy.

After the child had taken one of the toys to the chair and had played with it or examined it for a short time, he was requested to put it back into its original place. *E* then asked the child why he had chosen the one toy rather than the other. Answers given to this question were recorded, but the child was not compelled to reply. *E* then dismissed him with the toy he had been promised for this occasion.

As in Experiment IA, the differences in attractiveness of the two toys were counterbalanced by appropriate reversal of the conditions for each pair.

The experimenter in this procedure was C. W. S.

Results and Discussion

The results of this experiment confirmed those of Experiment IA. Of the 25 children, 19 chose to play with the toy which they were to receive 'in a few minutes,' while 6 chose the one to be given them in a week. The divergence form the hypothesis of equal distribution between the two alternatives was 2.6 times the standard error of the expected frequency, and it is therefore very unlikely to have occurred by chance. A χ^2 test of the homogeneity of the results of Experiments IA and IB, in the form of a two-by-two contingency table (with the use of Yates' correction for discontinuity), yielded a value of 1.125, for which the value of P is approximately 0.30. There is therefore no indication that the difference in the two procedures, including the use of verbal expression of preference in one as against behavioral choice in the other, had any significant effect upon the relative frequency with which the two alternatives were chosen. If the two sets of results are combined, it is seen that 41 out of 50 cases expressed a preference for the toy which was the sooner to be obtained. This frequency differs from that expected on the hypothesis of equal distribution by more than 5 times the standard error of the expected frequency. To the extent that the children's verbal and behavioral preferences constitute a valid criterion of the relative attractiveness of the toys, it may be concluded that the relative attractiveness of a toy is enhanced by its being made more nearly accessible in time.

These results appear to conflict with those of Wright. It must be noted, however, that Wright carefully distinguishes the effects of barriers which are difficult to pass, but nevertheless surmountable, from the effects of insuperable barriers. The latter he speaks of as 'field-limiting' factors; and he reports it as his experience that objects beyond such insuperable barriers tend simply to be ignored. "Thus, in three consecutive procedures, when the subjects . . . were asked to choose between a toy on a shelf far out of reach and another on a nearby table, after a single glance, the former was completely ignored" (**2**, p. 27). In the present experiments, it was observed that

the children paid more attention to the toy which they eventually chose, but they could not be said to have ignored the other. Nevertheless, it seemed advisable to determine the effect of a shorter time interval. Experiments IIA and IIB are concerned with this problem.

Experiment IIA

Subjects and Procedure

The 25 children who acted as subjects in this experiment had a mean age of 6.6 years. The age distribution was as follows:

Age	Frequency
9	1
8	5
7	9
6	2
5	8
Total	25

There were 9 boys and 16 girls. Eight of the children were colored and 17 white.

The toys used were similar to those of the previous experiments although not in general identical with them. Control of the possible inequalities in the original attractiveness of the toys was exerted by rotation of the toys within pairs, as before. Each toy was presented equally often on the right and on the left, except for the odd case of the 25, to determine whether bias was introduced by this positional factor. No significant bias appeared in the results, the numbers of preferences for toys on the right and on the left being 14 and 11, respectively. The instructions were as follows:

"Do you see these two toys? Well, I am going to give you this one [indicating it] in a few minutes, just as soon as I have finished writing here. After that I will start this little sand thing [indicating a three-minute sand-glass or 'egg-timer'], and when all that sand has run through like this [demonstrating the flow of sand], I will give you this toy [indicating the other toy]. You will be able to take them both with you when you go."

This was repeated until the child could reproduce it or answer correctly the questions, "When will you get this one [pointing]? When will you get this one [pointing to the other toy]? Which will you get first?"

The child was then asked, "Which toy do you like better?"

In an attempt to determine the relation between verbally and behaviorally expressed preferences, the procedure of Experiment IIA was extended as follows. After the child had stated his preference verbally, the experimenter said:

"Now, while I am writing here you may take either toy that you want to over to that chair there and play with it (or look at it). Now, remember, this toy [pointing] you will get when I am finished writing. Then I will start this little sand thing, and after all that sand has run through, you will get this other toy [pointing]. However, right now you may play with (or, look at) either one that you want over in that chair."

This experiment was carried out by F. M. A.

Results and Discussion

As can be seen in the procedure, this experiment introduced a period of only three minutes between the obtaining of the first toy and the obtaining of the second, instead of the one week's interval of Experiments IA and IB. The sand-glass was employed in order to make the length of the interval as concrete as possible for the child. In spite of these differences, however, 16 of the 25 children

expressed a preference for the 'immediate' toy as against the 'deferred' one. Although the difference from expectation here did not attain a high level of significance, being but 1.4 times the standard error of the expected frequency, it is consistent in direction not only with the results of the two experiments previously described, but also with those of the very similar Experiment IIB, as will be seen later.

All of the 16 children who stated verbally a preference for the immediate toy also chose that toy to play with under these conditions. Of the 9 children who stated verbally a preference for the deferred toy, 6 chose to play with this toy, but 3 chose to play with the immediate toy. This tendency for the two preferences to be the same is statistically significant. A value of P of less than .01 was obtained from a χ^2 test of the independence of the two sets of data, which means that there is significant dependence between the two sets of preferences. This does not, of course, show that the behavioral choices would have been the same if the previous verbal preferences had been eliminated. There is a possibility that the children felt a need to be consistent in their expressions of preference. The following experiment, however, demonstrated that the behavioral preferences, in the absence of verbal expression, favored the immediate toy when the three-minute interval was used.

Experiment IIB

Subjects and Procedure

The 25 subjects in this experiment were first-grade children in the Public School of Seaford, Delaware. Their age distribution was as follows: 1 at age 7, 22 at age 6, and 2 at age 5. The mean age was 6.0. There were 12 boys and 13 girls in the group. These were not the same children as those of Experiment IA, but were in a class for children retarded in reading proficiency as determined by scores on the Lee-Clarke Reading Readiness Test.

The general procedure was very similar to that of Experiment IIA. In the present case, however, no verbal preference was asked for. Otherwise, the instructions were identical with those of Experiment IIA, including the instructions requiring the child to take one of the toys to play with "while the experimenter was writing."

The experimenter was C. W. S.

Results and Discussion

As stated above, the results of this experiment were fully consistent with those of the three previous experiments. Of the 25 children, 18 chose to play with the immediate toy and 7 with the deferred toy. This result departs from chance expectation by 2.2 times the standard error of the expected frequency. If the verbal preferences of Experiment IIA are combined with these behavioral preferences, it is seen that 34 of the 50 children preferred the im-

mediate toy when the other was but three minutes more remote, a result which departs from expectation by 2.5 times the standard error. If, further, all differences in experimental procedure among the four experiments are neglected, it is found that 75 of the 100 children preferred the immediate toy. This outcome varies from chance expectation by just 5 times the standard error of the expected frequency. That it is justifiable to neglect the differences in experimental procedures in this way is indicated by the fact that a χ^2 test of the homogeneity of the four sets of preferences gives a value of P between .30 and .50. Since a test for the homogeneity of the results of Experiments IA and IB, combined, against the combined results of Experiments IIA and IIB, gave a value of P approximately equal to .10, there is little evidence that the difference between the one-week interval and the three-minute interval affected the preferences. Finally, combining the results of Experiments IA and IIA, and testing their homogeneity with the combined results of Experiments IB and IIB, gave a value of P equal to about .95, which indicates that the difference between the verbal and behavioral manners of expressing the preference had no significant effect upon the preferences.

Examples of replies which were given by the children to the question why they preferred one toy to the other are as follows:

"Because I wanted it."
"I love to play with them."
"I like to color." (coloring set)
"It looked prettier." (paper doll as against book)
"I want to wear it to the movies and church." (pin)
"Because I like to look at picture books."
"I liked this one best."
"I love cut-out books."
"I wanted to look at it." (air-raid warden set)
"Just 'cause I wanted to."

It is clear that these responses are either general, and not at all explanatory, or that they attribute the preference to some characteristic of the preferred toy. A few children (about 8 percent of those who were asked) gave answers which could be interpreted as referring in some way to the difference in temporal accessibility. The following serve as illustrations of such statements:

"Because I wanted to save this for a little later." (This in the case of a behavioral choice of the more remote toy)
"You told me I get this first, so I took it first." (In a case of behavioral choice of immediate toy)
"Because you're going to give me this one (*i.e.*, the remote toy) a week from now." (Behavioral choice of immediate toy)
"You're going to give this (the remote toy) to me next week. (Behavioral choice of immediate toy)
"You said I should take it first." (Behavioral choice of immediate toy)

Many of the children replied merely that they did not know why they preferred the one toy to the other. On the whole, the answers seem to throw little light upon the explanation of the preferences. They may, however, be taken to indicate that very few children were consciously aware of the influence of the time interval upon their preferences.

If the temporal intervals used in these experiments constituted barriers analogous to the spatial barriers studied by Wright, it must be concluded that our results are in apparent contradiction to his, and in general agreement with those of Fajans. This would be even more true in the comparison of our results with those of the experiments in which Wright used a temporal barrier; but we are inclined to believe that Wright's experiments with temporal barriers are more difficult to interpret than his others. At any rate, not only did the children in our experiments fail to prefer toys which were the more remote in time, but they clearly preferred the toys which were the nearer. This is hard to understand in terms of Wright's theory, which supposes that for an object to be almost in one's possession is equivalent to a partial enjoyment of it, and that this results in a diminution of its attractiveness. It is true that Wright reported greater difficulty in demonstrating the enhancement-value of a barrier in the case of children than in the case of adults; but this only weakens his own case, and could not account for results which are in the opposite direction.

We are not ready to offer a theory to explain the present results. However, it might be pointed out that there appear to be certain differences between temporal and spatial barriers which may have psychological meaning. A spatial distance can be overcome by expenditure of energy in bodily locomotion, and an interposed object may likewise be pushed aside or climbed over. A temporal interval, however, cannot in the same way be surmounted by active effort. One speaks of doing something 'to pass the time,' it is true; but whatever is done has the function of ridding oneself of the unpleasantness of a mere wait, and is not in itself a means toward attaining the desired object. It is possible that this difference is sufficiently fundamental, psychologically, to make it inappropriate to consider a time interval to be a *barrier* at all. Wright has defined barriers along a continuum determined by the extent to which they mediate or resist passage; and this we are inclined to accept. But, in the case of time, it can hardly be said that a short time mediates approach more than does a long time, since neither mediates approach in the same sense in which a physical path may. All time intervals are equally insuperable, so far as the efforts of the individual are concerned.

Summary

Two toys were presented before each of 100 children aged 4 to 9 years. The children were told that they would be given one of the two toys in the near future, and the other at a later time. The interval separating acquisition of the two toys was three minutes in the case of 50 children, and one week in the case of the remaining 50. From half of each of these two groups a verbal expression of preference between the toys was obtained; from the other half, a behavioral expression of preference was obtained. Of the 100 children, 75 expressed a preference for the toy which was the sooner to be acquired, and 25 preferred the more remote toy. No significant differences in the relative preference for the two toys were discovered to be associated with either the length of the time interval or the manner in which the preference was expressed. Very few (about 8 percent) of the reasons given by the subjects for their preferences referred to the difference in nearness of acquisition of the toys. The results suggest that time intervals do not function as barriers in a way analogous to spatial distances and interposed physical objects.

(Manuscript received April 19, 1943)

References

1. Fajans, S. Die Bedeutung der Entfernung für die Stärke eines Aufforderungscharakten beim Säugling und Kleinkind. *Psychol. Forsch.*, 1932, 17, 215–267.
2. Wright, H. F. The influence of barriers on the strength of motivation. *Contr. psychol. Theor.*, 1937, 1, No. 3. Pp. 143.

THE RELATIONSHIP BETWEEN ATTITUDES TOWARD CONCLUSIONS AND ERRORS IN JUDGING LOGICAL VALIDITY OF SYLLOGISMS [1]

BY IRVING L. JANIS AND FREDERICK FRICK

Introduction

This report deals with an experimental test of the efficacy of one of the factors which may underlie errors in reasoning. Woodworth and Sells (**1, 3**) have shown the operation of an 'atmosphere effect,' arising from the character of the premises, which leads to errors in syllogistic reasoning. Our interest is in determining whether or not the attitude toward the conclusion constitutes another factor which determines faulty judgment of the logical validity of concrete syllogisms.

It is a widely-held belief that people are likely to be satisfied with unsound arguments if they accept the conclusion to which the arguments lead, and conversely, that they are likely to be unduly critical of sound arguments if they reject the conclusion. The efficacy of this psychological fallacy is frequently assumed in discussions of the effects of propaganda and in evaluations of intellectual impartiality. In order to test this proposition experimentally, we restate it in the form of two hypotheses:

(1) If there is agreement with the conclusion, more errors will be made in judging logical validity by accepting invalid argumentation than by rejecting valid argumentation.

(2) If there is disagreement with the conclusion, more errors will be made in judging logical validity by rejecting valid argumentation than by accepting invalid argumentation.

Our experiment made use of syllogisms, and was designed to test two propositions which are subsumed under the two general hypotheses:

(1) There are more errors in judgments of the logical validity of syllogisms on 'agree-invalid' items than on 'agree-valid' items.

(2) There are more errors in judgments of logical validity of syllogisms on 'disagree-valid' items than on 'disagree-invalid' items. ('Agree' and 'disagree' refer to the attitude toward the conclusion

[1] This experiment was performed at Columbia University in the spring of 1941 as a project for Dr. Robert S. Woodworth's course in Advanced Experimental Psychology. The authors wish to express their gratitude to Dr. Woodworth for valuable suggestions and criticisms.

as an isolated statement; 'valid' and 'invalid' refer to the formal correctness of the reasoning leading up to the conclusion.)

The Syllogism Test

The two sub-hypotheses were tested by means of a list of 16 concrete syllogisms, 8 of which were logically valid and 8 logically invalid.[2] Syllogisms were selected whose conclusions were anticipated to produce either general agreement or general disagreement, with a minimum of neutral attitude responses. In this way it was possible to design the questionnaire so that there would be a fairly equal distribution of items in the following four categories:

(1) Agree-Valid
(2) Agree-Invalid
(3) Disagree-Valid
(4) Disagree-Invalid

The premises of the syllogisms in each of the four categories were examined with respect to the distribution of affirmatives, negatives, universals, and particulars in order to be sure that differential 'atmosphere effects' would not affect the results. Items in the four categories were presented in random order in the Syllogism Test. Definitions of the terms 'some' and 'all' were included in the instructions in order to avoid errors arising from the popular misconceptions of the meaning of these terms in logic.

The Syllogism Test was as follows:

Instructions: This is a test in reasoning. You are to check each of the following arguments as 'sound' or 'unsound.' A sound argument is one in which the conclusion follows logically from the premises. Do not concern yourself with the truth or falsity of the premises. (Note: 'All' in logic means *each and every* case. 'Some' means *at least one* and perhaps *all* cases.)

1.	No Bolsheviks are idealists and all Bolsheviks are Russian. Therefore, some Russians are not idealists.	
2.	The Eskimos are the only people who eat nothing but meat and it is found that all Eskimos have good teeth. So we may conclude that no people who eat only meat have bad teeth.	
3.	Many brightly colored snakes are poisonous. The copperhead snake is not brightly colored. So the copperhead is not a poisonous snake.	
4.	Some Russians are idealists. All Bolsheviks are Russians. It follows, therefore, that some Bolsheviks are idealists.	
5.	All poets die young but many professors are old, so we may conclude that not all professors are poets.	
6.	There is no doubt that some drugs are poisonous. All brands of beer contain the drug alcohol. Therefore, some brands of beer are poisonous.	
7.	All poisonous things are bitter. Arsenic is not bitter. Therefore, arsenic is not poisonous.	
8.	Some sailors are not able to swim. All Nantucketers can swim. Therefore, no Nantucketers are sailors.	
9.	No one has ever done any harm to his health by moderate indulgence in a mild drug. Nicotine is a mild drug and few cigarette smokers can afford to smoke excessively. Therefore, few people are harmed by cigarette smoking.	

[2] The items were selected from R. H. Thouless (2); in most cases, some modifications were introduced.

10. Man is ennobled only by the experience of finding himself faced by the choice of life
s and death. War provides the only situation in which men have to make this choice,
u so that, if universal and perpetual peace could be attained, it would be at the price of robbing men of all ennobling experience.
11. We must admit that the occurrence of quintuplets is an unusual event, but, to excite
s the legitimate interest of a reasonable person, it is not sufficient that an event should
u be merely unusual. So, no reasonable person should be interested in the Dionne quintuplets.
12. Many people fear that the prosperity of American industry may decline under increasing
s government control. They should remember, however, that the tariff system is itself
u a form of government control of industry and that our industries have prospered. This shows that government control is not harmful to industry.
13. No man can be blamed for any characteristic with which he was born or for any conse-
s quence of such a characteristic. Some men are criminals because they were born with
u criminal characteristics. So it follows that some criminals are not to be blamed for being criminals.
14. There are thousands who are waiting to emigrate from Germany when they get the
s chance. We may be sure that no one would wish to emigrate from a prosperous country,
u so we may conclude that any stories that we may hear of the prosperity of Germany are mythical.
15. It is found that houses in which a dog is kept are never visited by burglars. It is also
s found that houses with telephones are seldom visited by burglars. So it follows that
u some houses in which a dog is kept are fitted with telephones.
16. All good painting obeys the laws of perspective. Most of the paintings of Chinese
s and Japanese artists disregard these laws. Therefore, none of the paintings of Chinese
u or Japanese artists is of the highest rank.

Procedure

The Syllogism Test was administered to 19 subjects, all of whom were graduate students at Columbia University. None of the subjects had any training in formal logical analysis. There was no time limit on either the reasoning or the attitude test. After the reasoning test was completed, each subject was given the Attitude Test: the conclusions of all of the syllogisms were presented in random order and the subject was asked whether he disagreed or agreed with each of them.

Results

All items were tabulated into the four categories. The validity of the syllogisms was determined by the experimenters on the basis of logical rules; agreement and disagreement with the conclusion were determined empirically by the subjects' responses on the Attitude Test. The obtained distribution was converted into percentages (Table I) and used as the basis for the expected frequencies in a chi-square test. Each of the 19 subjects completed the list of 16 items, giving a group total of 304 responses on the reasoning test.[3] A total of 70 errors in reasoning was made by the 19 subjects. These errors were tabulated according to the four categories ('obtained freq.' in Table II). The error frequencies expected by chance ('expected freq.' in Table II) were calculated by multiplying the percentage of all items included in the given category (Table I) by 70,

[3] One item has been omitted from Table I, due to one subject's neutral response on the Attitude Test to one of the conclusions; hence the total is 303 responses rather than 304.

TABLE I

DISTRIBUTION OF ALL OF THE ITEMS JUDGED

Agree-Invalid	Disagree-Valid	Agree-Valid	Disagree-Invalid	Total
59 19.4%	83 27.4%	72 23.8%	89 29.4%	303 100%

TABLE II

CHI-SQUARE TABLE. DISTRIBUTION OF ERRORS

	Agree-Inv.	Disagree-V.	Agree-V.	Disagree-Inv.	Total
Expected frequency.....	13.6	19.2	16.6	20.6	70
Obtained frequency.....	24	22	13	11	70

Chi-Square = 13.6, $P < .01$.

the total number of errors. This gives the expected distribution of errors in the four categories, if the distribution was determined solely by chance.[4]

To determine whether there was a significant difference between the obtained distribution and that expected by chance, a chi-square test was applied. Chi-square was found to equal 13.6 and the corresponding probability value for three degrees of freedom is below .01, indicating that there is less than one chance in 100 of the deviation between expected and obtained frequencies arising solely by chance. This significant difference indicates that there is a greater tendency for errors to fall in certain of the categories than in others.

Do these results tend to confirm both of the sub-hypotheses or only one of them? In other words, does the tendency operate significantly for both cases of agreement with the conclusion and of disagreement with the conclusion, or is the significant difference found in Table II attributable to only one of these tendencies? To answer this question, the chi-square test was applied separately to the 37 items for which there was agreement with the conclusion and to the 33 items for which there was disagreement with the conclusion. The results appear in Tables III and IV. It is seen that the probability-values in both cases are below .05, indicating a significant difference for both tendencies.

These results tend to confirm the two sub-hypotheses on the relationship between attitude toward the conclusion and errors in the judgment of logical validity of syllogisms. Since this experiment

[4] Since group results are used—the total number of errors for all subjects combined—it was possible that the obtained distribution of errors was due to the results of only a few of the subjects. Examination of the data precluded this possibility, because errors were found to be fairly regularly distributed among all 19 subjects.

TABLE III

CHI-SQUARE TABLE. DISTRIBUTION OF ERRORS FOR ITEMS ON WHICH SUBJECTS AGREE WITH THE CONCLUSION

	Valid	Invalid	Total
Expected frequency	20.4	16.6	37
Obtained frequency	13	24	37

Chi-Square = 6.0, $P < .02$.

TABLE IV

CHI-SQUARE TABLE. DISTRIBUTION OF ERRORS FOR ITEMS ON WHICH SUBJECTS DISAGREE WITH THE CONCLUSION

	Valid	Invalid	Total
Expected frequency	15.9	17.1	33
Obtained frequency	22	11	33

Chi-Square = 4.5, $P < .05$.

was restricted to syllogisms, further tests on other types of material are necessary in order to increase the degree of confirmation of the two general hypotheses. It might be expected that tests of items in which the arguments are not so skeletonized will provide additional confirmation, because psychological factors affecting errors in the judgment of logical soundness probably are even more effective when it is difficult to discover the premises. Replications of this experiment with other types of subjects might also be expected to be confirmatory because graduate students are probably more cautious in their reasoning, particularly in minimizing bias, than is the population as a whole.

(Manuscript received April 28, 1943)

REFERENCES

1. SELLS, S. B. The atmosphere effect. *Arch. Psychol.*, 1936, No. 200.
2. THOULESS, R. H. *How to think straight.* N. Y.: Simon and Schuster, 1939. P. 246.
3. WOODWORTH, R. S., & SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. *J. exp. Psychol.*, 1935, 18, 451–460.

THE RETENTION OF THE ABILITY TO DRAW LINES OF A GIVEN LENGTH BLINDFOLD *

BY ELSIE O. BREGMAN, E. L. THORNDIKE, AND ELLA WOODYARD

We have studied the effect of disuse for varying periods of time upon a previously acquired ability—to draw lines 3, 4, 5 and 6 inches long with the eyes blindfolded. The subjects were three young women, a sophomore, a senior and a first-year graduate student.

All three subjects had received training in this performance, as follows: The blindfolded subject drew lines in columns upon a sheet of 16 × 21 inch cross-section paper, which was fastened by thumb tacks to a somewhat larger drawing board. Along the left edge of the board was affixed a thin strip of veneer. This served as a fixed starting place for all lines. The subject drew a line, starting at the strip, and immediately brought the pencil back to the strip, ready to draw another line. The length to be drawn was announced by the experimenter. The subject responded with a best approximation to that length, waited to hear if the attempt was scored 'right' or 'wrong,' and continued to draw the same length, waiting each time for the score, until directed to draw a different length. Three-inch lines were called 'right' if the error was not more than 1/8 in.; four-, five- and six-inch lines were called 'right' if the error was not greater than 1/4 in.

In the course of a single training period the subject drew 600 lines, 150 each of 3 in., 4 in., 5 in. and 6 in. The lengths followed each other in a random order, which was, however, constant throughout the experiment. No length occurred more than 8 or less than 4 times in direct succession.

The tests were like the training in length, content and procedure except that no announcements of 'right' or 'wrong' were made and that the experimenter named the length to be drawn for every trial.

Subject W was tested with 600 lines, then trained for 30 periods (18,000 lines), then tested after one day with 600 lines, then tested after 2, 4, 6, and 8 months. Before training W's average error (*i.e.*, deviation from the length she was asked to draw) was .82 in. for the 3-inch lines, .93 in. for the 4-inch lines, 1.27 in. for the 5-inch lines, and .98 in. for the 6 in. lines, a total of 4.90. Immediately

* The investigation reported here was one item in work of the Division of Psychology of the Institute of Educational Research of Teachers College, Columbia University, and was an incidental outcome of an investigation of the psychology of learning made possible by a grant from the Carnegie Corporation.

after the 30 periods of practice the corresponding amounts were .11, .13, .18, and .15, a total of .57. After 2 months disuse they were .12, .15, .14, and .18, a total of .59. After 4 months more of disuse they were .11, .14, .18, and .16, a total of .59. After 6 months of disuse they were .34, .33, .26, and .32, a total of 1.25, but on the evening of this test the subject became seriously ill. After 8 months of disuse, they were .12, .10, .11, and .12, a total of .45. On the whole, disuse up to 8 months seemed to leave W's ability unimpaired.

Subjects M and R had a test, 7 periods of training, a test, an interval of $3\frac{1}{2}$ months, 21 periods of training, a test, an interval of 2 months, a test, another interval of 2 months, and a test.

Their average errors in the test before the 7 periods of training were as follows (in inches) for 3-, 4-, 5-, and 6-inch lines in order:

M	.38	.54	.72	1.13;	total 2.77
R	.32	.40	.84	1.09;	total 2.65

After 7 periods of training (4200 lines) the corresponding amounts were:

M	.27	.23	.35	.36;	total 1.21
R	.33	.30	.40	.46;	total 1.49

After $3\frac{1}{2}$ months of disuse the corresponding amounts were:

M	.14	.19	.20	.23;	total .76
R	.16	.18	.21	.26;	total .81

After 21 more periods of training (12,600 lines) the corresponding amounts were:

M	.08	.09	.11	.11;	total .39
R	.09	.14	.15	.17;	total .55

Two months later with no further training the corresponding amounts were:

M	.23	.19	.26	.25;	total .93
R	.19	.33	.35	.39;	total 1.26

After another two months of disuse they were:

M	.11	.12	.15	.15;	total .53
R	.15	.19	.19	.27;	total .80

If we combine the 12 results for the four sorts of lines for the maximum length of disuse (8 months for W, 4 months for M and R) we find an average increase of 23 percent in the amount of error, with a probable error of 6. If we combine the 20 results for the four sets of lines for 4 months or $3\frac{1}{2}$ months of disuse, we find an average increase of 1 percent in the amount of error with a probable error of $6\frac{1}{2}$.

If we consider that the proper basis for computation is the three persons' total scores rather than the 12 or 20 scores for 3, 4, 5, and 6 inches, the average increases are 20 percent and 0 percent, and the probable errors are over twice as large. The most probable combined estimate is an increase in error of about 10 percent from 4 to 8 months of disuse.

It has been argued by some that the alleged consequences of an interval of disuse are really all caused by interference arising from other activities during the interval of disuse. This experiment adds to the evidence of deterioration in abilities by sheer disuse. There could have been little or no interference from other activities since the subjects were never in the experimental situation, and probably never in anything like it, during the intervals.

Interference is very important. We often forget one thing because we learn something else in place of it. The reason why the lapse of time weakens the connection between a certain situation and the response one has learned to make to it is very often that during that time the situation, or one more or less like it, has been connected with a different response and the response, or one more or less like it, has been connected with a different situation.

But we must also believe in a slow regression toward a sort of cerebral neutrality or zero, a slow weakening in connections uninterfered with save by the physiology of growth, decay, and other general changes. In senility this may outweigh interference from other activities; and as a limiting case we may have a person who does nothing but eat, sleep, and forget.

(Manuscript received April 13, 1943)